

MathSoft

**S-PLUS[®] FOR ARCVIEW GIS[®]
USER'S GUIDE**

Version 1.01

July 1998

Data Analysis Products Division
MathSoft, Inc.
Seattle, Washington

Proprietary Notice

MathSoft, Inc. owns both this software program and its documentation. Both the program and documentation are copyrighted with all rights reserved by MathSoft. No part of this publication may be produced, transmitted, transcribed, stored in a retrieval system, or translated into any language in any form without the written permission of MathSoft, Inc.

See the License Agreement and Limited Warranty for complete information.

The correct bibliographical reference for this document is as follows:

S-PLUS for ArcView GIS User's Guide, Data Analysis Products Division, MathSoft, Seattle, WA.

Copyright Notice

Copyright © 1987-1998, MathSoft, Inc. All rights reserved.
MathSoft, Inc.
101 Main Street
Cambridge, MA 02142
USA

Printed in the United States of America.

Trademarks

S-PLUS is a registered trademark, and STATSERVER, S+SDK, S+SPATIALSTATS, S+DOX, S+WAVELETS, and AXUM are trademarks of MathSoft, Inc.

S and New S are trademarks of Lucent Technologies, Inc.

ArcView GIS and ARC/INFO are registered trademarks of Environmental Systems Research Institute, Inc.

Intel is a registered trademark and 486, SX and Pentium are trademarks of Intel Corporation.

Microsoft, Windows, MS-DOS, and Excel are registered trademarks and Windows NT is a trademark of Microsoft Corporation.

Other brand and product names referred to are trademarks or registered trademarks of their respective owners.

Acknowledgment

MathSoft gratefully acknowledges the substantial support from NASA Earth Observations Commercialization Application Program (EOCAP'93) contract NAS13-623 for our project entitled "*SpaceStat: An Object-Oriented Software Toolkit for Statistical Analysis of Spatial and Spatial/Temporal Data*". We thank the Project Liaison Support personnel, particularly Paul Maughan and Brett Thomasie, for their support during the course of this project.

TABLE OF CONTENTS

1	Welcome	1
	Installation	2
	What You Can Do with S-PLUS for ArcView	3
	Tips on Learning S-PLUS for ArcView	4
	Technical Support	6
2	Quick Start Tutorial	7
	Exercise 1: Exploring Demographic Data	9
	Exercise 2: Finding Relationships Between Demographic Variables	18
	Exercise 3: Geographic Clustering of Linear Regression Modeling Errors	24
3	Using S-PLUS for ArcView GIS	31
	Overview	32
	The S-PLUS Menu	34
	The Spatial Statistics Menu	46
	The Help System	54
4	Using S-PLUS and S+SPATIALSTATS	55
	What is S-PLUS	56
	Object Browser	56
	Graph Sheets	65
	Data Windows	69
	What is S+SPATIALSTATS	79
	Script Window and Commands Window	81

5 Analyzing Spatial Data in S-PLUS for ArcView	85
Spatial Statistics	86
Data Preparation	86
Defining Neighbor Relationships	89
Spatial Autocorrelation	91
Local Spatial Association	94
Spatial Regression	98
Bibliography	103
Index	105

WELCOME

1

Installation	2
What You Can Do with S-PLUS for ArcView	3
Tips on Learning S-PLUS for ArcView	4
Typographic Conventions	5
Technical Support	6

Welcome to S-PLUS for ArcView GIS. This software package combines desktop GIS and mapping with powerful statistical and graphing tools for visualizing, exploring, and analyzing spatially referenced data. S-PLUS for ArcView consists of:

- ArcView GIS
- S-PLUS
- S-PLUS for ArcView extension
- S+SPATIALSTATS add-on module (optional)

S-PLUS for ArcView connects ArcView with the S-PLUS data analysis software through a number of menu-driven dialogs. You will work with S-PLUS for ArcView through ArcView's **S-PLUS** and **Spatial Statistics** menus, shown below.



Figure 1.1: *The S-PLUS and Spatial Statistics menus in ArcView.*

This manual explains how to use S-PLUS for ArcView and gives a brief introduction to both S-PLUS and S+SPATIALSTATS. This chapter covers:

- how to install S-PLUS for ArcView
- what you can do with S-PLUS for ArcView
- how to learn S-PLUS for ArcView.

Installation

S-PLUS for ArcView requires S-PLUS Version 4.5 or later. To install the S-PLUS software, refer to the instructions that come with the software. It is a good idea to turn off other applications, in particular virus checkers, while installing S-PLUS.

To install the S-PLUS for ArcView extension:

1. Insert the CD-ROM into your CD-ROM drive.
2. From the system Explorer, double-click on the **setup.bat** file in the root of the CD-ROM.

What You Can Do with S-PLUS for ArcView

In ArcView, you create maps which incorporate geographic data, and summarize this data in a flexible, visual environment. With S-PLUS for ArcView you can now penetrate your data with unprecedented power, using the data analysis tools in S-PLUS. What is more, you can graphically explore the geographic data using the cutting-edge, intuitive S-PLUS graphing tools, and integrate both graphics and analysis results into the ArcView layout.

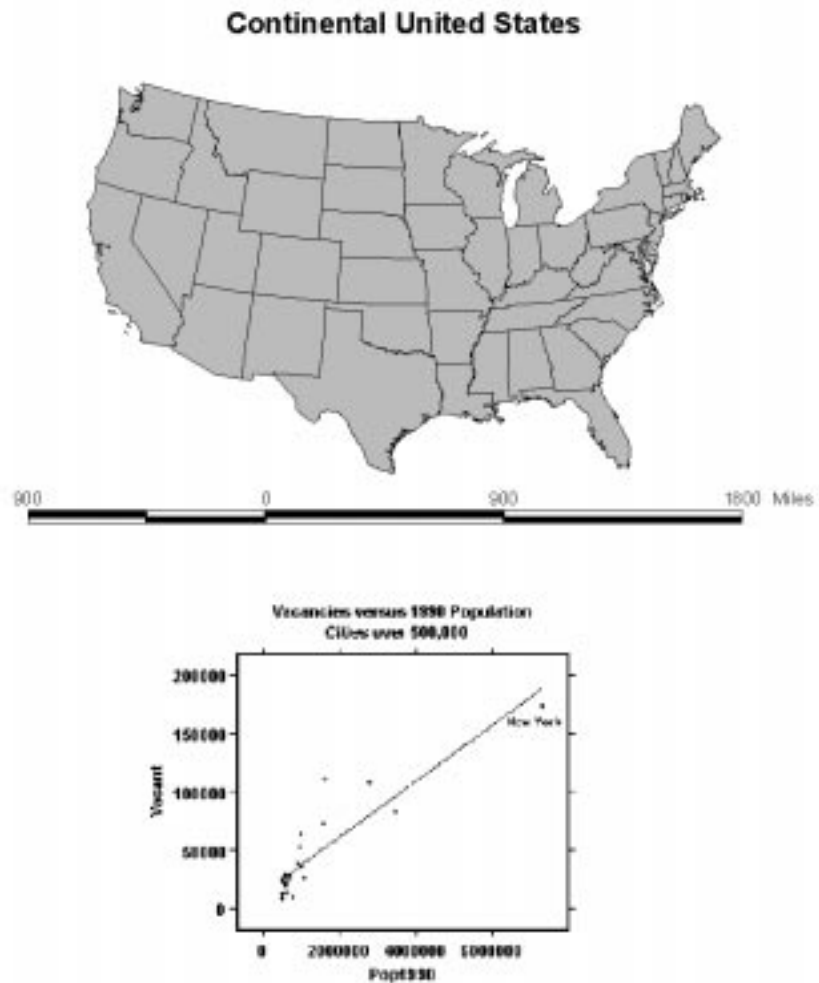


Figure 1.2: Simple linear regression output from S-PLUS in an ArcView layout.

In S-PLUS for ArcView, you can access S-PLUS directly from ArcView to analyze data in the attribute tables of themes. Other tables may be exported to S-PLUS, and the results later joined in ArcView to theme attribute tables. See the release notes for details on which data types are currently supported.

Tips on Learning S-PLUS for ArcView

- This manual focuses in turn on S-PLUS for ArcView, S-PLUS, and S+SPATIALSTATS. A good place to start is in chapter 2, *Quick Start Tutorial*. Starting with some typical data in ArcView, the exercises illustrate concretely how you can extend your analysis and visualization with S-PLUS for ArcView. The exercises are designed so you can work along with them on your computer.

The next chapter, *Using S-PLUS for ArcView GIS*, gives detailed descriptions of the functionality in the **S-PLUS** and **Spatial Statistics** menus in ArcView. You may want to skim chapter 3 at first reading, to take just a brief look at the ArcView interface to S-PLUS. This chapter will be a helpful reference later on when you face specific tasks.

You'll want to work in S-PLUS to create graphs of your multi-dimensional GIS data, and to calculate summary statistics or fit models based on geographical data. Go to chapter 4, *Using S-PLUS and S+SPATIALSTATS*, for a look at the S-PLUS environment, and for an overview of how to work in S-PLUS and its S+SPATIALSTATS module.

In chapter 5 you will see in more how to use the S+SPATIALSTATS module in S-PLUS to analyze spatial data.

- **When running S-PLUS for ArcView:** choose **S-PLUS Language Reference** in the Help menu to open the S-PLUS online help system. In the help index you can access help files for over 2000 S-PLUS functions.

Choose S-PLUS for ArcView Help for information on any of the items under the **S-PLUS** and **Spatial Statistics** menus in ArcView.

This manual can also be accessed in electronic format from the Help menu.

- **When running S-PLUS:** choose **Using S-PLUS** in the Help menu for information on the S-PLUS user interface.

Choose **Questions and Answers** for a discussion of some common difficulties, and their solutions.

The **Visual Demo** item launches an online demo to help users familiarize themselves with common tasks in S-PLUS.

The *S-PLUS User's Guide*, the *S-PLUS Programmer's Guide*, and the *S-PLUS Guide to Statistics* can all be accessed in electronic form from the S-PLUS Help menu.

- **In both ArcView and S-Plus:** you can access context-sensitive help by clicking the Help button in any dialog, or by clicking the context-sensitive Help button  in the main toolbar.
- When working in the S-PLUS Commands window, type `help()` at the S-PLUS prompt, or press F1, to open the S-PLUS online help system.

To display the index of S+SPATIALSTATS help files, use the S+SPATIALSTATS Help item in the ArcView Help menu.

Typographic Conventions

This manual follows these typographic conventions and references:

- The *italic* font is used for emphasis, and also for the names of S-PLUS manuals.
- The **bold** font is used for filenames.
- The typewriter font is used for S-PLUS commands and output. Displayed S-PLUS commands follow the S-PLUS prompt `>`. Commands that require more than one line of input are displayed with the S-PLUS continuation prompt `+`.
- Key names appear in capitalized form. For example, the shift key appears as Shift.
- When more than one key must be pressed simultaneously, the two key names will appear with a plus sign between them. For example, the key combination of Shift and F1 appears as Shift+F1.
- The arrow keys are useful for moving objects around the page. They appear as Up, Down, Left, and Right.

Technical Support

Here is how to contact S-PLUS technical support:

North America

voice: **(206) 283-8802 x230**, Monday through Friday, 7:30 a.m. to 5:00 p.m. Pacific Time.

fax: **(206) 283-6310**

e-mail: **support@statsci.com**

Europe, Asia, Australia, Africa and South America

voice: **+44 1276 452299**

fax: **+44 1276 451224**

e-mail: **shelp@mathsoft.co.uk**

QUICK START TUTORIAL

2

Exercise 1: Exploring Demographic Data	9
Display the Atlanta View	9
Compute Summary Statistics in ArcView and Prepare Data for Export	10
Export the Cleaned Attribute Table to S-PLUS	11
Locate the Table in S-PLUS	12
Make Histograms Comparing Percentages of Infants	13
Import an S-PLUS Histogram into an ArcView Layout	15
Exercise 2: Finding Relationships Between Demographic Variables	18
Plot the Relationships Between Three Fields	18
Find the Relationship Between Average Income and Percentage 0 to 4 Years, Conditioned on Percentage Females	19
Visualize Average Income, Percentage 0 to 4 Years, and Percentage Females	22
Exercise 3: Geographic Clustering of Linear Regression Modeling Errors	24
Fit a Least Squares Linear Regression Model	24
Color Classify the Census Tracts Using Residual Values	26
Add a Spatial Barchart Based on Residuals	27
Improve the Spatial Barchart Using Positive and Negative Residuals	28

The S-PLUS scientific and statistical graphing capabilities provide a comprehensive complement to the basic charting capabilities in ArcView. With the S-PLUS graphing tools you will discover more valuable information in your ArcView data, summarize it more effectively, and produce presentation-quality charts.

This chapter contains three detailed lessons on extending your work in ArcView with S-PLUS for ArcView. They are designed so that you can walk through each step along with the exercise. In order to do the exercises, you will need to have installed S-PLUS and S-PLUS for ArcView on your system. It is not necessary to have installed the S+SPATIAL-STATS module to do these exercises.

EXERCISE 1: EXPLORING DEMOGRAPHIC DATA

In this exercise you will find out how to:

- Clean up the demographic data in the attribute table for the Atlanta census tracts
- Export this clean version of the data to S-PLUS
- Generate histograms for several of the fields for informative comparisons
- Generate scatterplots of pairs of fields with automatic linear and nonlinear curve fits
- Import a graph back into an ArcView layout

Display the Atlanta View

1. *Start ArcView*, typically by selecting Program from the Start Button, then navigating to ESRI and ArcView GIS.
2. *Open the qstart.apr project.* From the File menu, use “Open Project...” to navigate to and then open the `\ESRI\AV_GIS30\AVTutor\ArcView\qstart.apr` project.
3. *Start the S-PLUS for ArcView extension.* With the **qstart.apr** window in focus, select the Extension dialog, from the File menu. In the dialog, check the S-PLUS for ArcView GIS item in the list of available extensions. When OK is pressed to confirm your selection, the extension will load. If S-PLUS is already running, the extension connects to S-PLUS; otherwise, S-PLUS will be started which may take a minute or two.
4. *Open the Atlanta view.* Once the extension has loaded, the window titled **qstart.apr** will display several views. Select the Atlanta view and then press the Open button in the window.
5. *Open a theme table.* Click on the “Census.tracts” theme to enable most tools and many menu items. Click the Open

Theme Table button  on the ArcView toolbar. The attribute table of the Census tracts theme opens. This table has 482 records and 47 fields, with one record per census tract. The fields of this table are demographic attributes such as average income, median income, population in 1990, and average age.

Shape	City	State	Tract	Sq. miles
Polygon	057	090200	057090200	65
Polygon	057	090100	057090100	68
Polygon	117	130200	117130200	46
Polygon	117	130100	117130100	36
Polygon	057	090900	057090900	68
Polygon	057	090400	057090400	17
Polygon	117	130300	117130300	36
Polygon	117	130400	117130400	29
Polygon	057	090300	057090300	42
Polygon	117	130500	117130500	34
Polygon	057	090600	057090600	22

Figure 2.1: The attribute table of the Census Tracts theme.

Compute Summary Statistics in ArcView and Prepare Data for Export

Before exporting your table to S-PLUS to create graphs, it is a good idea to check for anomalous data which should not be used in making graphs. Here is how to perform such a check for the Atlanta census tract demographic data.

6. *Clean the data.* Select the 25th field, **Avg_inc**, a field of interest. Choose Statistics from the Field menu. This displays summary statistics in a pop-up window.

Notice that the minimum value of the average income is -99, which does not make sense. The numerical value -99 has been used to denote missing data. Now you need to find out how many such values there are in the table.

To do this quickly, first choose Sort Ascending from the Field menu, with the field **Avg_inc** still selected. The top portion of the table now appears as in figure 2.2.

The first four rows of the table have entries of -99 in all fields. Evidently, no data was available for these census tracts. The fifth row is rather suspicious because it has many zero values. From the sixth row on, the data look reasonable.

Select rows 6 through 482 as follows. First select the first 5 rows by clicking each of them while holding down the shift key. Then click the Switch Selection button  on the ArcView toolbar.

You are now ready to export the cleaned data to S-PLUS.

Avg_0_50	Avg_inc	Avg_50_100	Avg_100+
-99.000000	-99.000000	-99.000000	-99.000000
-99.000000	-99.000000	-99.000000	-99.000000
-99.000000	-99.000000	-99.000000	-99.000000
-99.000000	-99.000000	-99.000000	-99.000000
0.000000	0.000000	0.000000	0.000000
100.000000	2500.000000	2500.000000	1428.561
44.900000	9532.549900	5434.779900	2636.841
57.759999	10496.890000	4308.509900	4760.561
69.139999	13505.290000	3613.230000	6538.271
66.899999	13941.790000	4104.790000	7222.021

Figure 2.2: The attribute table of the Census tracts theme, after sorting on the Avg_inc field.

Export the Cleaned Attribute Table to S-PLUS

7. Open the Export dialog. Click the title bar of the Atlanta view to make it active. Now choose Export Data to S-PLUS from the S-PLUS menu. The dialog shown in figure 2.3 appears.

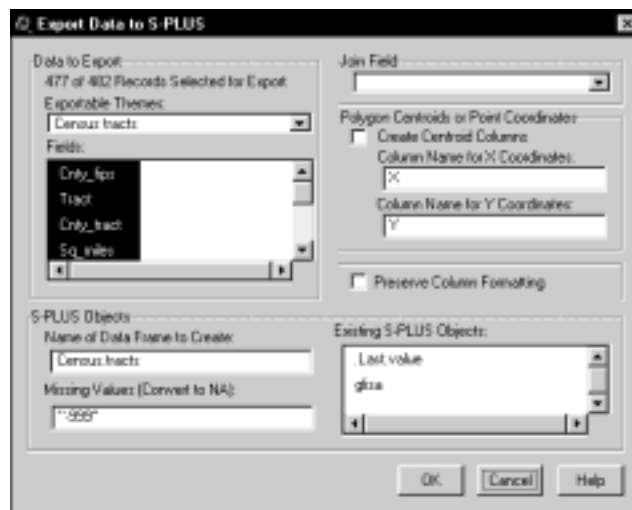


Figure 2.3: The Export Data to S-PLUS dialog.

This dialog allows you to save an ArcView table as an S-PLUS data frame. An S-PLUS data frame is a table of data which can have numeric, character, and other kinds of data in its fields. The dialog reports that only 477 of 482 records are selected for export, as desired.

The glisa function is created each time that S-Plus for ArcView starts up.

S-Plus names must start with a letter, and can contain only letters, numbers, and periods. Under-scores, in particular, are not permitted.

Join fields are used by the extension any time new tables or fields are to be joined (displayed together with) to the theme's original attribute table.

8. *Name the new data frame.* By default, the name of the S-PLUS data frame to be created is based on the ArcView theme name. There is no risk of overwriting the ArcView table, since the S-PLUS data frame is saved in a different directory. Keep the name that is displayed below Name of Data Frame to Create or change it to another name. Scroll through the list below Existing S-PLUS Objects to see all the objects in the S-PLUS “working directory.” If you are using S-PLUS for the first time this working directory has the name **_Data**, and contains only the objects `.Last.value` and `glisa`.

If you have already been using S-PLUS there may be several additional objects in your working directory. By default, the dialog displays a data frame name that does not exist in S-PLUS. If you specify an existing object's name, a warning dialog will confirm your choice.

9. *Select a Join Field.* Next, select Tract from the list of fields in the Join Field combo box. This identifies a field containing unique values for each record and is helpful if you ever need to re-import the data or analysis results from S-PLUS back into ArcView. Click OK to export the Census tracts demographic data to S-PLUS.

Locate the Table in S-Plus

10. *Look in the Export dialog.* Open the Export to S-PLUS dialog again, and scroll down the list below Existing S-PLUS Objects. In the list you will see the name of the data frame just created.

When you loaded the S-PLUS for ArcView extension, S-PLUS was started automatically. Move to S-PLUS now by selecting it from the Task Bar. You can confirm that the data frame `Census.tracts` matches the selection in the Census tracts attribute table in ArcView.

If the Object Browser button is depressed and the Object Browser is not visible, choose Object Browser from the Window menu to bring it forward.

11. *Look in the Object Browser.* If an S-PLUS Object Browser window is not open, depress the Object Browser button  on the toolbar. The Object Browser works in a way similar to the Windows Explorer. In the left-hand pane, click the plus sign next to the data.frame icon to expand the list. In the list below it is `Census.tracts`.

12. Look at the data in a Data window. Double-click the Census.tracts icon to open a Data window, as shown in figure 2.4. It contains the 477 records exported from the Census tracts attribute table.

	22	23	24	25	26	27	28
	Inc 5.15k	Inc 0.5k	Avg inc	Med inc	Per cap inc	Pct male	Pct female
1	11.97	3.87	43698.59	39074.86	15859.11	50.14	49.86
2	14.79	8.00	38468.02	32680.36	14171.58	49.78	50.22
3	11.66	5.06	46647.83	39540.00	17177.46	49.78	50.22
4	13.03	3.54	43074.34	36199.18	16455.45	50.87	49.13
5	10.18	6.46	45232.62	40809.02	16563.27	50.55	49.45
6	28.12	0.81	33017.46	28052.12	12691.10	47.43	52.57
7	9.68	4.48	47750.89	41073.57	17255.66	49.67	50.33

Figure 2.4: The exported data in an S-PLUS data frame.

Data windows in S-PLUS are similar to table windows in ArcView. See chapter 4, Using S-PLUS and S+SPATIALSTATS, for an overview, and chapter 4 of the *S-PLUS User's Guide* for complete details on what you can do in the S-PLUS Data window.

Make Histograms Comparing Percentages of Infants

Barplots, such as those available in ArcView, are ideal for *categorical* data with relatively few categories. Such data sometimes arise from grouping numeric data into intervals.

Histograms, on the other hand, are ideal for numeric data that take on a *continuous range* of values. For example, the percentages of infants aged 0 to 4 years in each of the Atlanta Census tracts data are well suited to visualization with a histogram, because the percentages vary over a continuous range of values.

Here is how to construct histograms in S-PLUS using numeric data exported from ArcView.

13. *Select the data to plot.* Click the S-PLUS Object Browser window to make it active. Expand the data.frame icon in the left-hand pane, if necessary, by clicking the plus sign. The Census.tracts icon is displayed in the left pane; click it. In the right-hand pane you will see icons with names corresponding to the fields of the attribute table of the Census tracts theme. Select Pct.0.4yrs by clicking on it.

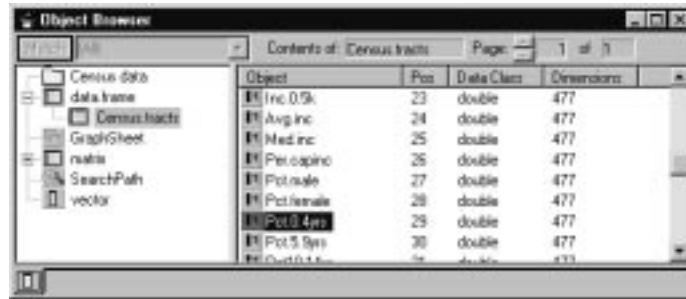


Figure 2.5: Selecting the Pct.0.4yrs column in a data frame using the Object Browser.

14. Create a histogram. Depress the 2D Plots button  on the S-PLUS toolbar to make the Plots2D palette appear. Click the Histogram button  on this palette. A graph sheet appears with the histogram shown in figure 2.6. The histogram is

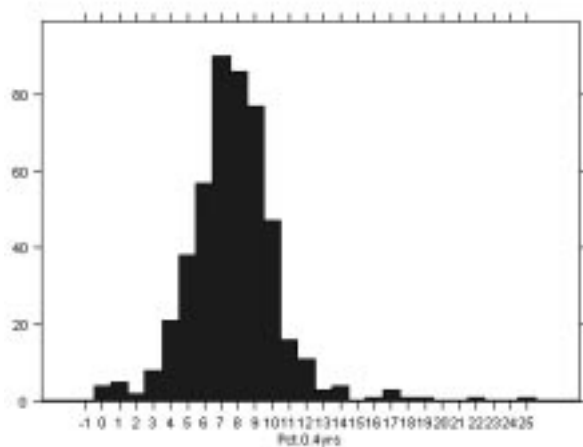


Figure 2.6: Histogram of the percentages of infants aged 0 to 4 years in the selected Atlanta census tracts.

essentially unimodal, with a longer tail on the right. We immediately notice some rather high values; some census tracts report over 20% of the population being between 0 and 4 years old. There is a cluster of census tracts for which the percentage of the population in this age range is under 2%.

The number of classes and the class size for the above histogram has been computed using an automatic algorithm, based on our 477 values of Pct.0.4yrs. In this case 26 classes resulted.

S-PLUS allows nearly every feature of a graph to be customized for technical presentation purposes. Right click on various parts of the graph to see the various property dialogs available.

15. *Change the title and axis label.* To add annotations clarifying the data being displayed, first double click the X axis title and change it from the default (@AUTO) to “Percent of Census Tract Population Under 4 Years Old” and then, from the Insert menu, select Titles, Main..., and enter “Infant Distribution in Atlanta Census Tracts” to set the main title.
16. *Change the number of classes.* To change the number of classes, first click one of the histogram bars. You clicked in the right place if a green handle appears below the first histogram bar. Right-click on the histogram, and choose Options from the context menu to bring up the Histogram/Density property dialog. Change the entry beside Number of Bars from Auto to 50, and click OK. The histogram is redrawn with 50 bars instead of 26, which shows finer detail.

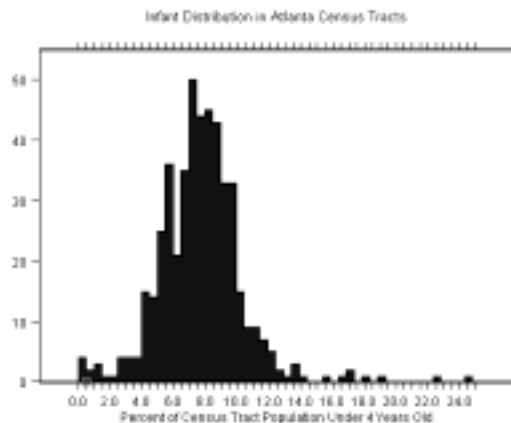


Figure 2.7: Histogram of the percentages of children aged 0 to 4 years in the selected Atlanta census tracts, using 50 classes.

Import an S-PLUS Histogram into an ArcView Layout

17. *Import the graph.* Switch back to ArcView, by selecting it from the Windows Task Bar. Then, with the Atlanta view open and selected, use ArcView's S-PLUS menu to open the Import S-PLUS Graph dialog. If you do not have an existing layout, you

will be asked whether to create one. Specify Yes. The resulting dialog, shown in figure 2.8, allows selection from a number of plots open in S-PLUS, the file and format used to store the graphic, and the destination layout in ArcView. Since the defaults suffice, simply click the OK button to actually import the graph from S-PLUS.

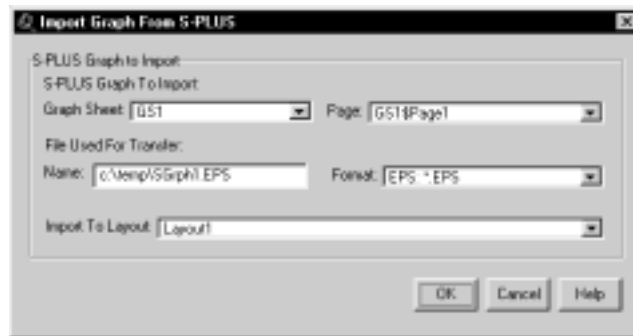


Figure 2.8: *The Import Graph From S-PLUS dialog.*

18. *Display and format the graph.* Display the imported graph by selecting Layouts in the qstart.apr project window in ArcView, then selecting the newly created Layout1, and pressing the Open button. Detailed graph features may not display in the layout but will print fine. See figure 2.9.
19. *Closing the extension.* When you are done using the extension, go to the project window, qstart.apr in this example, so it has focus, and then under the File menu, select the Extensions dialog and uncheck the S-PLUS for ArcView GIS extension.

As for all ArcView extensions, if you close and save the project while an extension is active, that extension will automatically be loaded the next time you load the project.

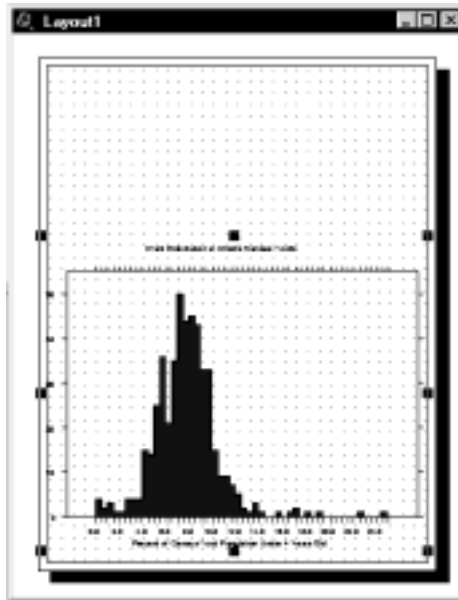


Figure 2.9: An ArcView layout with an S-PLUS graph sheet.

EXERCISE 2: FINDING RELATIONSHIPS BETWEEN DEMOGRAPHIC VARIABLES

In this exercise you'll look some more at the Atlanta census demographic data. In particular, you'll do the following with tools in S-PLUS:

- Visualize the relationships between three or more variables using pairwise scatterplots
- Visualize the relationships between three or more variables using interactive brush and spin plots
- Visualize the relationships between several variables using the unique Trellis graphics displays of conditional views of the data. We will focus on the relationships between Average Income and Percentage 0 to 4 Years within groups of records classified by Percent Female.

This exercise takes up where Exercise 1 leaves off. If you have not already walked through Exercise 1, perform steps 1 through 9 from that exercise now to create the data set that is used here.

Plot the Relationships Between Three Fields

1. *Select the fields.* In the S-PLUS Object Browser, click on the data frame `Census.tracts` in the left-hand pane. Its fields will be shown in the right-hand pane. Then Ctrl-click `Avg.inc`, `Pct.female`, and `Pct.0.4yrs`, in that order. All three fields will be selected.
2. *Create a scatter plot matrix.* Depress the 2D Plots button  to open the Plots 2D palette. Click the Scatter Plot Matrix button . The plot in figure 2.11 appears in a Graph sheet.

Several of the plots show features outside the main cluster of points. The plot of Average Income versus Percentage 0 to 4 Years shows a negative association, and other plots show a small number of census tracts with extreme values of Percentage Female.

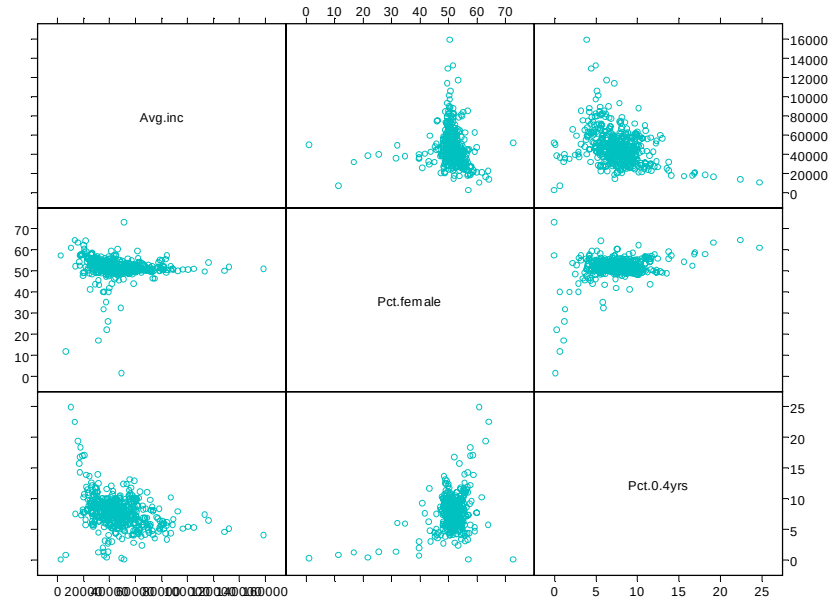


Figure 2.10: A scatterplot matrix showing several pairwise relationships.

Find the Relationship Between Average Income and Percentage 0 to 4 Years, Conditioned on Percentage Females

You can discover detailed relationships between several fields by *conditioning* on one of the fields. This means looking at the relationship between two fields on various *subsets* of the data set. The subsets are determined by the values of a third field, in much the same way that you use a classification field to specify a graduated color legend in Arc-View.

3. Create a scatter plot with a loess curve fit. For example, first make a scatterplot of Avg.inc versus Pct.0.4yrs with a loess nonlinear curve fit. Ctrl-click Avg.inc and then Pct.0.4yrs in the Object browser, and click the Loess button  on the Plots 2D palette. The plot shown in figure 2.11 appears.

The scatterplot is the same as that in the lower left panel of figure 2.10, except for the addition of the loess fit. The loess fit shows little nonlinearity, and does not fit the points in the upper left of the plot. To make a tighter fit, right click the plotted points and choose Smooth/Sort from the menu which appears.

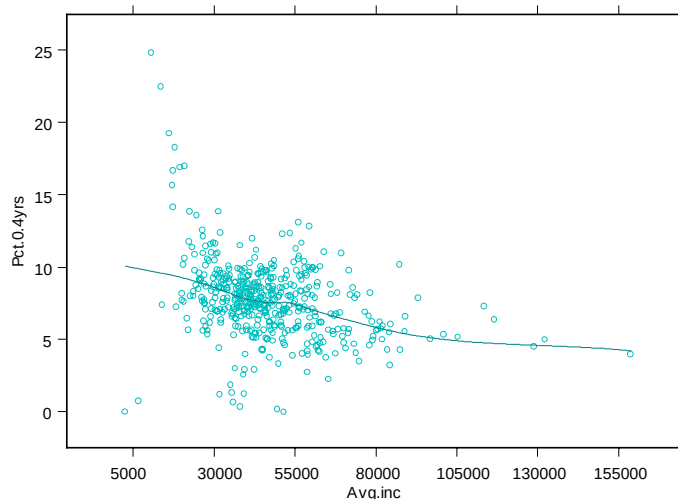


Figure 2.11: Atlanta census tract Average income versus Percent Population 0 to 4 Years, with a loess curve fit displayed.

Next to Span, enter 0.1 and click OK. The plot shown in figure 2.12 appears.

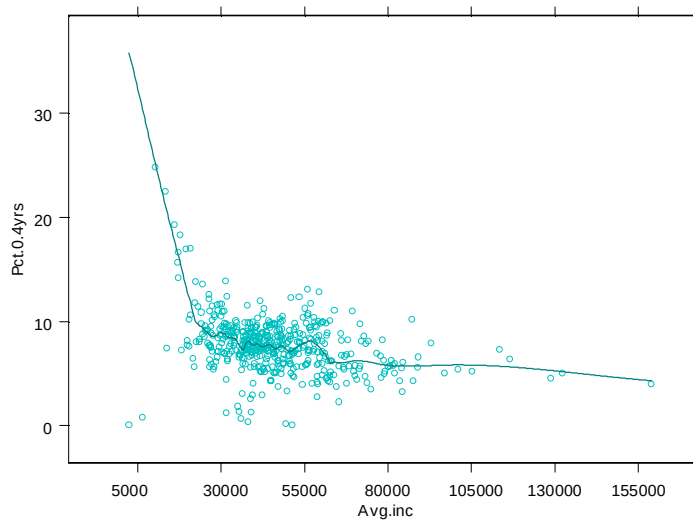


Figure 2.12: Atlanta census tract Average income versus Percent Population 0 to 4 Years, with a loess curve using a span of 0.1.

How does this relationship depend upon Pct.female? To explore this, you will use Pct.female as a *conditioning* field to divide the records into several classes, and then display the relationships in all the different classes in a single S-PLUS Trellis plot.

4. *Turn on conditioning mode.* First, turn on the conditioning mode for graphics by depressing the Set Conditioning Mode toolbar button . Be sure that the number of modes, seen just beside, is .
5. *Select the fields.* Ctrl-click Avg.inc, Pct.0.4yrs, and Pct.female in this order.
6. *Create the plot.* Click the Loess button , which now carries a horizontal bar to indicate that conditioning mode is on. The plot shown in figure 2.13 appears.

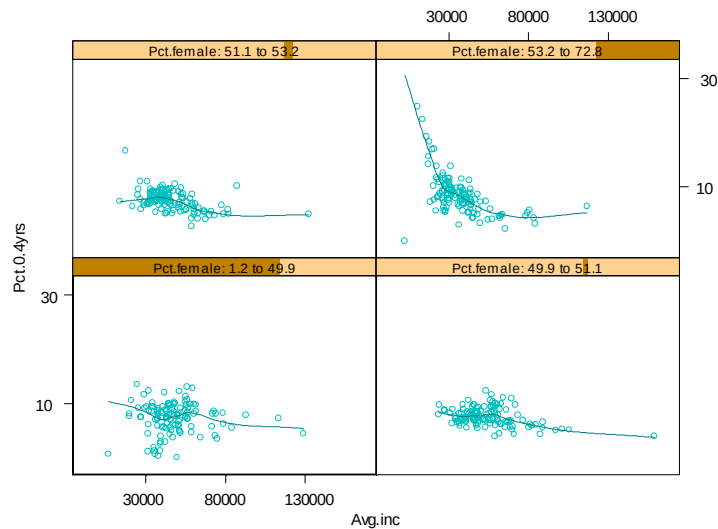


Figure 2.13: Trellis plot showing the relationships between Average Income and Percentage Population 0 to 4 Years in 4 groups of census tracts. The groups are classified by Percentage Female.

This Trellis display clearly reveals relationships which are not easily seen otherwise.

- For census tracts with the highest percentage of females, there is the strongest negative association of average income and percentage 0 to 4 year-olds.
- Virtually all the census tracts with a high percentage of 0 to 4 year-olds fall among those which have over 53.2 percent females.
- There is somewhat greater variation in the percentages of 0 to 4 year-olds among those census tracts whose populations are unbalanced with respect to gender.

Visualize Average Income, Percentage 0 to 4 Years, and Percentage Females

To explore these data interactively, S-PLUS provides a graph window with tools for rotating a 3-dimensional scatterplot and selecting subsets of data simultaneously in several scatterplots.

7. *Select the fields.* Choose Brush and Spin from the Graph menu. The Brush and Spin dialog prompts you for a data frame and some of its fields, as seen in figure 2.14.



Figure 2.14: *The Brush and Spin dialog.*

Hold down Ctrl while selecting the fields in order to make a multiple selection.

8. *Explore the 3D data interactively.* Click the arrow buttons in the upper right of the Spin window to rotate the scatterplot of Avg.inc, Pct.0.4yrs, and Pct.female. You can see the 3-dimensional features of the plot.

9. Sweep a selection of the data. Drag the pointer across any of the plots, and the points which it covers are highlighted in all plots and in the list of data points in the lower right. Since the box next to Persistence is checked, the points remain highlighted after the pointer passes over them.

To leave the Brush and Spin windows, click the Quit button.

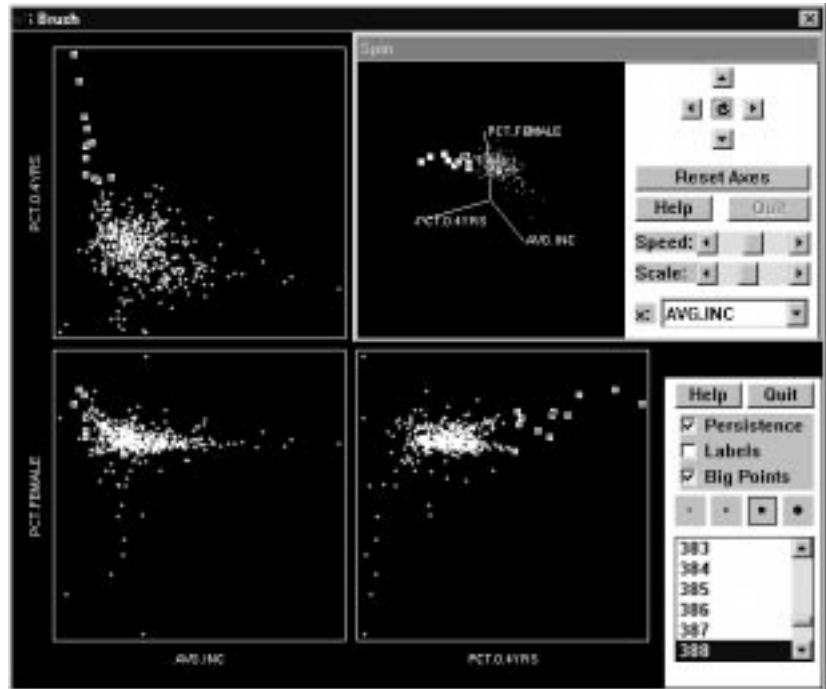


Figure 2.15: S-PLUS Brush and Spin windows.

10. Print a view of the interactive plot. To print a snapshot of the interactive Brush and Spin windows, it is necessary to copy the screen image to the clipboard, and then save the image using an image editing program. Press the Print Screen button on your keyboard to copy the screen image. Then open Microsoft Paint, or Microsoft PowerPoint and paste into a blank document. Save the file. You can then import it into an ArcView layout.

EXERCISE 3: GEOGRAPHIC CLUSTERING OF LINEAR REGRESSION MODELING ERRORS

In these exercises you will fit a simple linear regression model for predicting average income based on average age and percentage male for each of the Atlanta census tracts. You will visualize the model errors, or residuals, for each census tract using both color classification and spatial bar charts. The exercise follows these steps:

Caution: Use ordinary least squares linear regression with spatial data only as an exploratory tool. To make correct conclusions about spatially correlated data, use spatial linear regression, available from the Spatial Statistics menu in ArcView when the S+SPATIALSTATS module for S-PLUS has been installed.

- Fit a Linear Least Squares Regression Model
- Color Classify the Census Tracts According to Residuals Values
- Add a Spatial Barchart Based on Residuals
- Improve the Barchart by Creating and Using Positive and Negative Residuals

Fit a Least Squares Linear Regression Model

1. *Select the data.* Follow steps 1 through 6 of Exercise 1. This selects all but 5 rows of the census tract data.
2. *Display Linear Regression Dialog.* Select the Atlanta view and ensure the Census tracts theme is selected to enable the items in ArcView's S-PLUS menu. Select Linear Regression... from the menu and accept the warning dialog indicating that the Linear Regression will be computed only on the 477 selected records.
3. *Specify the model.* In the resulting dialog, figure 2.16, select the fields Avg_inc, Pct_male, Avg_age from the first column in that order. This generates the formula $\text{Avg\_inc} \sim \text{Pct\_male} + \text{Avg\_age}$ indicating that Avg_inc is the response to be modeled as a function of Pct_male and Avg_age. Specify Tract as the Join Field, if it is not already set. Finally ensure that "Display File after Computation" is checked, and click OK.



Figure 2.16: *The Linear Regression dialog.*

Caution: Avoid the temptation to believe the *t*-statistics and *p*-values. Reliable values for these quantities are given by a spatial linear regression model.

The fields selected will be exported to S-PLUS for analysis, creating the Census.tracts.lm data frame in S-PLUS. Once done, results of the linear regression such as coefficient values, *t*-statistics, and *p*-values are displayed a window. Upon closing this window, the dialog closes. The resulting Linear Fits and Residuals are imported to a new table, “Linear Regression Analysis of Census.tracts.lm”, and then joined to the “Attributes of Census tracts” table. See figure 2.17.

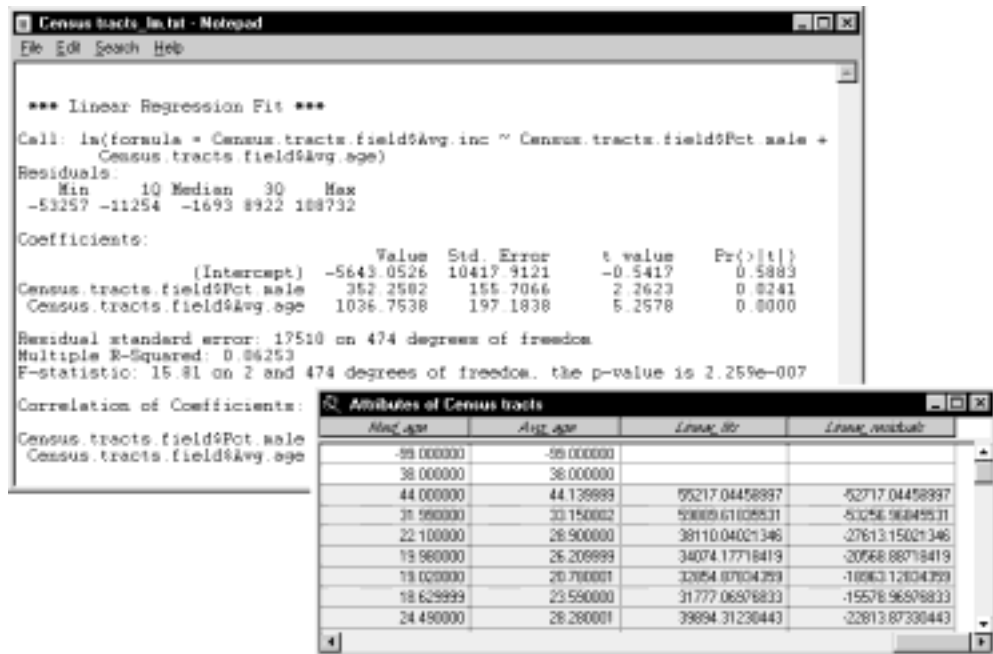


Figure 2.17: Linear Regression Model results and selected record set.

Color Classify the Census Tracts Using Residual Values

4. *Prepare the view.* With the View window in focus, click the Zoom to Full Extent button  on the ArcView toolbar. Turn off the Highways theme, but keep the Streets in downtown theme turned on to provide a visual indicator of the downtown census tracts. From ArcView's S-PLUS menu, select Clear Selection and Graphics to unselect the various census tracts.
5. *Color classify the Census tracts.* Select Color Classification from the S-PLUS menu. In the dialog that appears, click the Theme button, and for the Census tracts theme select the last numerical column, Last_res. Move the Color Classes slider to 8, and click OK.

The Census tracts now appear color-coded, with the legend indicating coding according to the value of the residuals in the linear regression model.

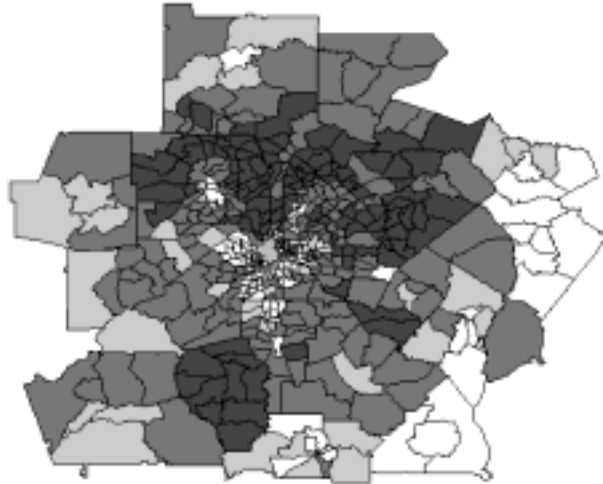


Figure 2.18: *Atlanta census tracts after color classification by residuals.*

Some geographic patterns are now clear. The most positive residuals (model errors) tend to cluster in groups to the north of the downtown area, and in a few areas to the south of downtown. The most negative residuals tend to cluster in the downtown and the region just to the south and west of downtown, and in groups near the boundary of the region.

In other words, the simple linear regression model which predicts average income with average age and percent male has errors whose values are geographically clustered, or correlated. The clustering may suggest how to improve the simple linear predictive model by adding other predictor variables.

Add a Spatial Barchart Based on Residuals

6. *Remove the color classification.* Change the color classification in the previous step to a Single Symbol Legend type with a symbol having white foreground.
7. *Create spatial barcharts.* Click the View window to make it active, and choose Spatial Bar/Pie Chart from the S-PLUS menu. In the dialog that appears, select Census tracts in the list below Theme, and select Last_res as the Field. Then select Blue

below For Each Column Select a Color. At this point, {Last_res, Blue} appears in the Command box. Under Chart Properties, specify Bar Chart and change the scale factor default value to 0.75. Click OK.

The result, shown in figure 2.19, gives a good indication of the location of the model errors having largest magnitude, and gives a similar sense of clustering of large model errors. Barcharts provide a better visual comparison of the relative size of model errors than our color classification did.



Figure 2.19: *The Atlanta Census tracts with Spatial bar charts.*

Improve the Spatial Barchart Using Positive and Negative Residuals

8. *Identify positive and negative residuals.* To visualize more readily the residuals we will classify them by sign. You will now add two fields to the Census tracts theme, one containing only the positive valued residuals, with zeros elsewhere, and one containing only the negative valued residuals, with zeros elsewhere. Begin by selecting Tables on the qstart.apr project window. Double click on Attributes of Census tracts to open this table. From the Table menu, select Start Editing, then under the Edit menu bring up the Add Field dialog twice and create two new fields, res.positive and res.negative as shown in figure 2.20.

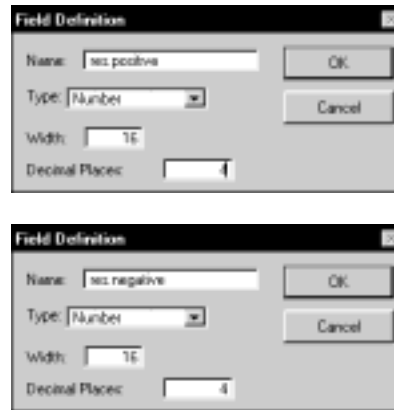


Figure 2.20: *Field Definition dialogs in ArcView.*

Select the positive residuals by opening the Query dialog by clicking on the Query Builder button . In sequence, double-click the [Linear_residuals] field, the greater or equal  button, and then 0 from the keyboard. Finally select the New Set button to select just those records that meet this criteria and close the dialog.

To insert the positive residuals into the new res.positive field, click once on the field name at the top of the table and then bring up the Calculate dialog via the Field menu.

Then reverse the selection, using the Switch Selection button , select the negative residuals field in the attribute table, and use the calculate dialog again to insert the negative residuals into the res.negative field. When done, the Attributes of Census tracts table should look like figure 2.21

9. *Create bar charts.* Now you will make spatial barplots, with red bars for negative residuals and blue bars for positive residuals. Click the View window to make it active, and choose Spatial Bar/Pie Chart from the S-PLUS menu. In the dialog that appears, select Census tracts under Theme, and select res.negative under Field. Then select the color Red. You will see {res.negative, Red} in the Command box. Go again to the Field list, and select res.positive. Select the color Blue. You will see

avg_age	linear_ov	linear_residual	res.positive	res.negative
34.439999	46893.14314345	-3194.55314345		-3194.5531
35.790001	47485.26727220	-9017.36727220		-9017.3673
34.419998	46899.12910898	-251.30110898		-251.3011
34.869999	47059.03467057	-3985.03467057		-3985.0347
34.990000	47123.22629219	-1890.59929219		-1890.5993
36.750000	48065.17474410	-15047.71374410		-15047.7137
33.430000	46450.32523965	1300.56576435	1300.5658	
35.119999	47251.29434419	-2992.04534419		-2992.0453
33.509998	46393.98142111	-8632.05142111		-8632.0514
35.919998	47547.86132794	9054.21667206	9054.2167	

Figure 2.21: Census tracts attribute table showing new fields.

{res.negative, Red} + {res.positive, Blue} in the Command box. Under Chart Properties, specify Bar Chart and change the scale factor to 0.75. Click OK.

- The result, shown in figure 2.22 below, gives a clear visual indication of where the large positive model errors are clustered and where the large negative model errors are clustered.

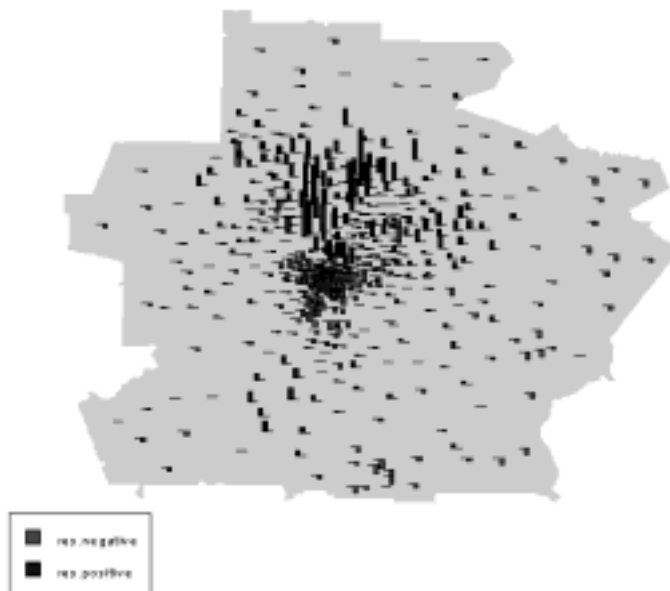


Figure 2.22: Atlanta census tracts showing spatial bar charts for both positive and negative linear model residuals.

USING S-PLUS FOR ARCVIEW GIS

3

Overview	32
Running S-PLUS for ArcView	32
The S-PLUS Menu	34
Export Data to S-PLUS	34
Import Table from S-PLUS	36
Import Point Theme from S-PLUS	37
Group Selected Records	37
Add Indicator Field	39
Color Classification	40
Spatial Bar/Pie Chart	41
Import Graph from S-PLUS	42
Clear Selection and Graphics	43
Linear Regression	43
Execute S-PLUS Commands	44
The Spatial Statistics Menu	46
Spatial Neighbors	46
Spatial Autocorrelation	49
Local Spatial Association	50
Spatial Linear Regression	52
The Help System	54
S-PLUS for ArcView Help	54
S-PLUS for ArcView User's Manual	54
S-PLUS Language Reference	54
S-PLUS Spatial Statistics Help	54
About S-PLUS for ArcView	54

OVERVIEW

S-PLUS for ArcView GIS, an extension to ESRI's ArcView GIS program, allows users to access the powerful statistical capabilities of MathSoft's S-PLUS program. S-PLUS for ArcView does more than provide a convenient mechanism for moving data and graphics between the two programs. It brings the vast analytical and graphical functionality of S-PLUS into ArcView. Further, S-PLUS for ArcView is fully integrated with the optional S-PLUS Spatial Statistics module. When that module is installed with S-PLUS, ArcView users have direct access to a variety of sophisticated spatial statistical analysis and visualization tools. Use S-PLUS for ArcView to

- export selected records and fields from ArcView tables to S-PLUS data frame objects
- import S-PLUS data sets to ArcView tables or point themes
- record row selections by storing character or numeric values in a new or existing field
- color classify a theme based on data from S-PLUS or ArcView
- create graphic bar or pie charts relating data on a per polygon basis
- import S-PLUS graph to an ArcView layout
- clear graphics and selections
- calculate linear regression fits and residuals for selected records
- execute S-PLUS commands directly.

In addition, if you have the optional S+SPATIALSTATS module installed, you can:

- create neighbor objects, also known as spatial weight matrices
- calculate global (Spatial Autocorrelation) or localized (Local Spatial Association) measures of spatial autocorrelation
- estimate spatial regression models.

Running S-PLUS for ArcView

To run S-PLUS for ArcView, follow these simple steps.

1. Start ArcView on your computer.
2. From the File menu, select Extension.

3. Choose S-PLUS for ArcView GIS from the list of available extensions.
4. Click OK. **S-PLUS** and **Spatial Statistics** menu items appear in the main menu bar, as in figure 3.1, indicating that S-PLUS for ArcView is active. Note that most menu items are only enabled



Figure 3.1: S-PLUS and Spatial Statistics appear in the main menu.

if a view document is in focus, and has a theme selected which is based on a shape file.

THE S-PLUS MENU

Use the **S-PLUS** menu to exchange data with S-PLUS, to place bar charts or pie charts on a view, or to use S-PLUS to analyze theme data. You can execute single S-PLUS commands on theme data, or use S-PLUS to perform analysis and visualization of theme data.



Figure 3.2: *The S-PLUS menu.*

Export Data to S-PLUS

Select Export Data to S-PLUS to export selected fields and records into a new S-PLUS table, called a data frame. This dialog can also be used to add centroid information to the attribute table. To export data, follow these steps:

1. Choose a theme from the drop down list under Exportable Themes.
2. Select the fields to be exported from the ArcView attribute table associated with the chosen theme. These fields will be exported into the data frame. To select more than one field, hold down the SHIFT key while clicking.
3. If the data is to be imported back into ArcView and joined to an existing table, you must specify a field to be used as a Join Field. The field specified must have all non-duplicate values. A check for record uniqueness is performed at selection time
4. Specific geographical information of spatial observations is often required when conducting a spatial analysis. The x- and y-coordinates for a point coverage or the centroids of a polygonal coverage are typically useful in spatial analysis. To create centroid columns in the exported data frame, check the

box under Polygon Centroids or Point Coordinates and specify names for the x- and y-coordinate columns, such as Latitude and Longitude, for instance. You must specify a Join Field when centroid coordinates are requested.

5. Specify a name for the S-PLUS data frame to be created. By default a unique name is suggested.
6. Check whether the resulting columns in the S-PLUS data frames should preserve the same formatting (width and precision) as in the theme's attribute table. Choosing this option will increase the time needed to export the data considerably.
7. If there are missing value codes, such as -999, in the data, enter these values below Missing Values (Convert to NA). Separate multiple values with a comma. Surround strings with double quotes. These will be converted to the special NA (Not Available) value when the data is transferred to S-PLUS.

The selected fields and records are extracted from the shape files, and are transferred into S-PLUS as an S-PLUS data frame, with the specified name.

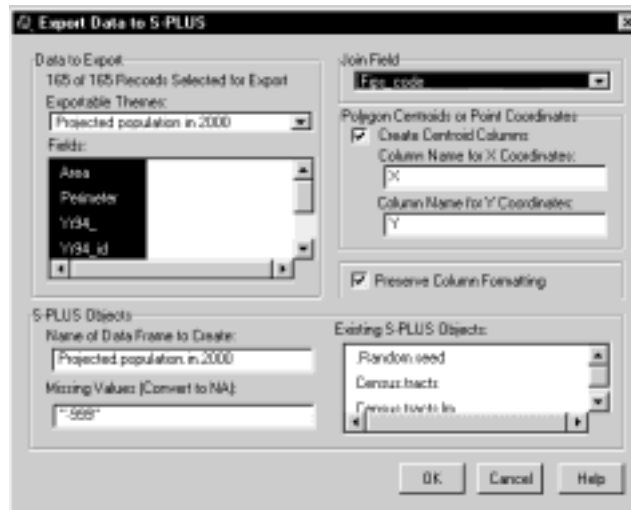


Figure 3.3: *The Export Data to S-PLUS dialog.*

Import Table from S-PLUS

Use this dialog to import an S-PLUS data frame as a table in the current project. The variables from the data frame can then be displayed in the View window, if joined to an existing table. For example, you can plot a residual or fitted value for each polygon in a map.

1. Select a data frame from the list of S-PLUS objects.
2. Select the columns of the S-PLUS data frame to import. To select more than one column, hold down **SHIFT** while clicking.
3. Specify a value to denote missing values in the imported table. Instances of NA in the S-PLUS data frame will be converted to this value when imported to ArcView.
4. Specify a name for the imported table.
5. To join the fields in the imported table to the attribute table of a theme, click the checkbox next to Join Imported Columns to Theme, and choose a theme from the drop down list.
6. Specify the names for the common field in each table under Match Records Based on Common Values....
7. Click OK. The new table is added to the list of Tables in the current ArcView Project window.

Joining two tables causes them to be temporarily merged.

Joining the imported table to an existing table requires one field from each table with corresponding values for each record to be joined.



Figure 3.4: The Import Table From S-PLUS dialog.

Import Point Theme from S-PLUS

Use this menu item to convert an S-PLUS data frame into a point theme and add it to the ArcView View window. The S-PLUS data frame must contain X and Y coordinates as column variables. These can have any names, such as Longitude and Latitude. The selected S-PLUS data frame is imported into ArcView as a new table.

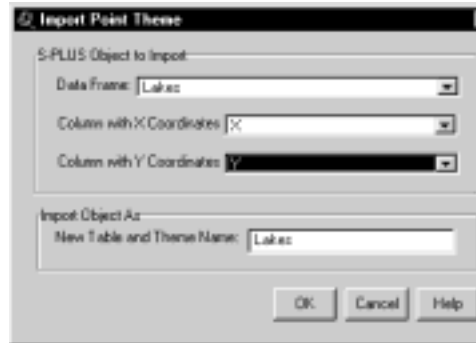


Figure 3.5: *The Import Point Theme From S-PLUS dialog.*

1. Select an S-PLUS data frame from the drop-down list.
2. Select the variables which contain the X and Y coordinates in the data frame. These must be different columns.
3. Specify a name for the new theme and its table.
4. Click OK. The added table is converted into a point theme and is shown in the ArcView View window.

Group Selected Records

Use this dialog to create a new character field that records which records are selected. Once one or more selections have been recorded, the table can be exported to S-PLUS and analyses done to spot trends for each of the groups selected.

Use the Group Records dialog when you wish to denote the selection with a character value and use the similar Add Indicator Field dialog when you wish to denote the selection with a numeric value. These dialogs record the selection in either a new field created on the spot or in an existing field.

Before opening the dialogs, select some features in a theme. Some tools for this include the query dialog, custom Avenue scripts, or the selection tool. If no features were selected, all records will be included in the selection. To group the records into numerous categories, repeatedly make a new selection then call the dialog.

1. Use the radio buttons on the left to specify whether to create a new field or overwrite values of an existing field.
2. If you write to a new field, you will need to specify a Join Field, so that the new field can be joined to the active theme's attribute table.
3. If you write to an existing field, only values in currently selected rows will be overwritten. In this case, you can choose either to keep or to clear values in non-selected records. The text at the top of the dialog indicates how many of the total records are currently selected.
4. Specify a character value (a category name or group identifier) to write to each of the records in the current selection. Click OK.

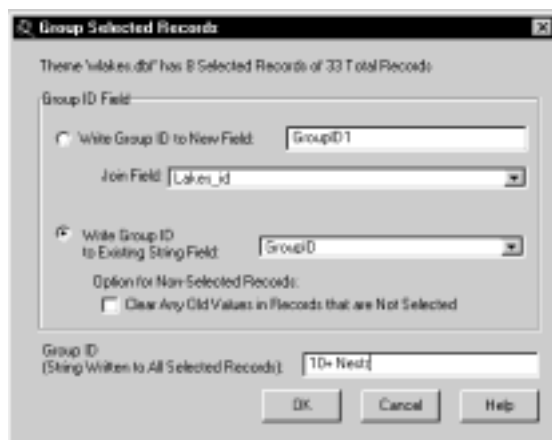


Figure 3.6: *The Group Selected Records dialog.*

For example, select counties on the North side of the state and group ID those using the string “North”; then reverse the selection, and group the newly selected records by writing to the same field and using the string “South” as the Group ID. If you export the resulting field to S-PLUS it will result in a categorical variable (a factor) with two levels.

Add Indicator Field

Similar to Group Selected Records, the Add Indicator Field dialog adds a numeric field that records which records are selected. This is useful after selecting a subset from the attribute table of the active theme. An indicator field can be added to the attribute table for labeling the selected subset of records. This field can then be used in S-PLUS to subset corresponding variables or as a dummy variable in linear regression models or other statistical analyses.

Before opening the dialogs, select some features in a theme. Some tools for this include the query dialog, custom Avenue scripts, or the selection tool. If no features were selected, all records will be included in the selection. To group the records into numerous categories, repeatedly make a new selection then call the dialog.

1. To create a new field, check the corresponding button and specify a name for the indicator variable. Specify a Join Field, so that the new field can be joined to the active theme's attribute table.
2. To write the indicator values to an existing numeric field, check the corresponding button and choose a field from the drop-down list. For selected records, choose whether to replace current values of this field with the indicator value, or to add the indicator value to the current value. For non-selected records, choose whether to clear or to keep the current values.
3. Specify an integer value to write to each of the records in the current selection. Click OK.

For example, select counties where the level of a chemical has risen above some threshold and assign an indicator value of 1 to those records; then once exported to S-PLUS you might evaluate records above the threshold with those below.



Figure 3.7: *The Add Indicator Field dialog.*

Color Classification

This menu item allows you to use color to categorize features in the active theme by values in any numeric field. For example, you may color states based on their population levels.

1. Select a theme or an S-PLUS data frame from one of the two drop-down boxes. The theme or data frame must have the same number of records as the active theme; its data must be associated with the same region of study. If a data frame is selected, then it must have been exported from ArcView with a Join Field from the active theme.
2. After selecting the theme or data frame, its fields (columns) will be displayed. Select one field from the list of fields.
3. Specify the number of quantiles or classes by using the slider. The range is from 1 to 25.
4. Click OK. The selected variable is displayed in a map in the View window. The color scale is automatically determined by the requested number of classes.
5. A legend shows the corresponding ranges of values for each color class. These ranges can be edited individually to create classes that are differently sized. The labels can also be edited. Double-click the legend to open the legend editor.

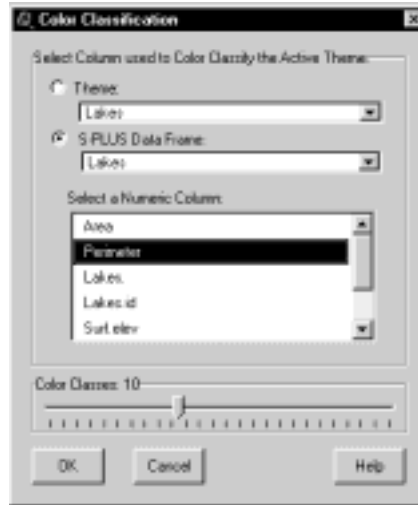


Figure 3.8: *The Color Classification dialog.*

Spatial Bar/Pie Chart

This function plots a bar or pie chart using data from fields in the current theme or in any theme with the same number of records as the active theme, that is, with data associated with the same region of study.

1. Build a plotting command from fields in the selected theme and colors for the corresponding bar or pie slice. Select a theme and a field. The field name appears in the command line.
2. Select a color with which to plot the attribute selected above. You can also type directly in the command field. Repeat steps 1 and 2 as desired for additional bars or slices.
3. Choose between bar chart and pie chart, and specify a scale factor for the chart. Specifying a number less than 1 makes a smaller chart; specifying a number greater than 1 makes a larger chart.
4. Click OK to generate the chart.
5. Use the Clear Selection and Graphics choice from the S-PLUS menu to remove the plots from the current view.

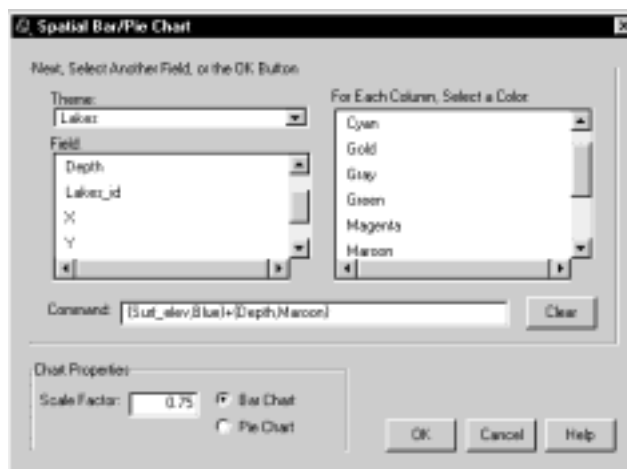


Figure 3.9: *The Spatial Bar/Pie Chart dialog.*

Import Graph from S-PLUS

To transfer a graph created in another application into ArcView, you would normally export the graph as a file in one of the file formats supported by ArcView, open a layout in ArcView, draw a picture frame on the display, and place the graphics file in the picture frame. The Import Graph From S-PLUS dialog automates this procedure.

If no layout currently exists, you are first prompted to create a new one. Graphs in S-PLUS are placed on Graph Sheets which can contain multiple pages.

1. Select an S-PLUS Graph Sheet from the drop down menu. Plot names cannot contain spaces.
2. If the Graph Sheet has more than one page, select the desired page from the list.
3. Choose a format for the graphics file with which the plot will be exported from S-PLUS and imported into ArcView.
4. Specify an alternate name for the graphics file to be created if so desired.
5. Choose a layout into which to place the plot.
6. Click OK.

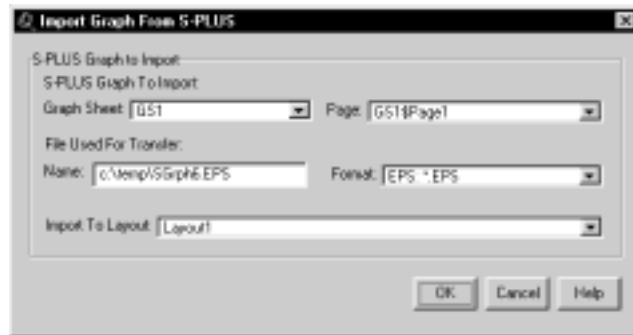


Figure 3.10: *The Import Graph from S-PLUS dialog.*

Clear Selection and Graphics

Choose this item after drawing the Spatial Charts on the spatial units in the View window. It resets the picture to its state before the charts were added. When a table is selected, this menu item will clear the selection in the same way as the Clear Selection toolbar button.

Linear Regression

Use this menu item to fit a linear regression model based on variables selected from the attribute table for the active theme or from S-PLUS data frames.

1. Build a formula from fields in the current theme and columns in S-PLUS data frames in the S-PLUS data directory. Click a theme field or data frame column to add it to the formula. The first selection made is used as the response variable. You cannot type directly in the formula field.
2. If desired, choose a vector of weights from among the columns of the currently selected data frame or theme. The default is to fit a model without regard to weights.
3. An S-PLUS data frame containing fitted and residual values will be saved after the regression runs. Specify a name for this object and indicate whether you would like to import it as a table and join it to the current theme for further exploration in ArcView.
4. A summary report of the regression will be saved in a text file. Specify a name for the report file and indicate whether it should be opened immediately after running the regression. After

reviewing the summary information as displayed, make sure to close the displayed file to complete the operation and close the dialog.



Figure 3.11: *The Linear Regression dialog.*

The OK button remains disabled until a formula with two variables at least has been selected and a join field in the case of a requested join.

Execute S-PLUS Commands

This menu item provides easy access to S-PLUS commands from ArcView. Use this to run analytical routines or perform simple operations without having to invoke the S-PLUS user interface.

1. Refer to the S-PLUS Objects list for the names of existing objects in the S-PLUS data directory. Choose an object class, such as data.frame, to view only that type of object. When a data frame is selected, its columns are displayed to the right.
2. Use the buttons below the Objects list to browse, copy, rename, or delete an S-PLUS object in the S-PLUS data directory before issuing any command.

3. Compose an S-PLUS command in the top field. Press Enter or click on the Run button to issue the command. If you do more work in the dialog after entering the command, click first on the command line and then press Enter to issue the command. The output appears in the S-PLUS Output window.
4. Use the checkboxes below the command line to paste a copy of each command either to the clipboard or to the S-PLUS output window in the dialog.



Figure 3.12: The Execute S-PLUS Commands dialog

THE SPATIAL STATISTICS MENU

For users who have access to the S+SPATIALSTATS module of S-PLUS, S-PLUS for ArcView takes advantage of the ease of identification of geo-referenced data in ArcView. S-PLUS for ArcView facilitates the computation of spatial neighbor relationships used in the computation of spatial autocorrelation, spatial regression, and other spatial statistics. As with the S-PLUS menu, a view must be in focus to have access to this menu.

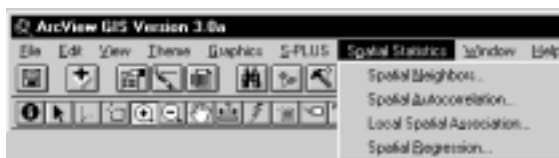


Figure 3.13: *The Spatial Statistics menu.*

Spatial Neighbors This dialog is designed to construct spatial weight objects for spatial statistics and modeling. The concept of a spatial weight is important in spatial statistics and spatial modeling. Many criteria have been suggested for constructing various forms of spatial weights, and different weighting schemes may significantly affect the results of a spatial analysis.

In S-PLUS and S+SPATIALSTATS, there is an object structure called `spatial.neighbor` for managing spatial weights. S-PLUS also provides a `read.neighbor` function to read geographical neighbor information from a text file. However, the user must construct the text file containing the spatial neighbor information. This can be time-consuming for a study region having many spatial units.

S-PLUS for ArcView provides an easy tool for constructing spatial neighbors (spatial weights) objects. The spatial weights are calculated based upon the geo-locational information included in ArcView shape files. The Spatial Neighbors dialog allows the user to select among several options for calculation of the spatial weights.

First Order Neighbor Weights

This option constructs a spatial weight matrix based solely on the adjacency of spatial units (polygons). The element $x[i, j]$ of the resulting weight matrix

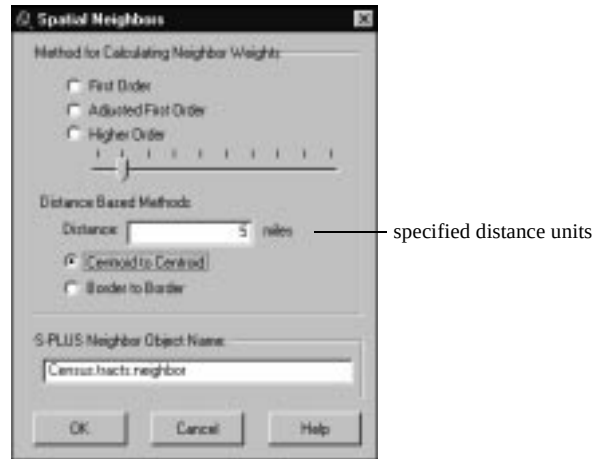


Figure 3.14: *The Spatial Neighbors dialog.*

X is 1 if polygon j is adjacent to polygon i , and is 0 otherwise. The diagonal elements $x[i, i]$ of the spatial weight matrix X are all equal to 0.

Adjusted First Order Neighbor Weights

In most cases, polygons of a theme are irregularly shaped. A polygon may be a short distance from a given polygon i without sharing a boundary with it. The simple spatial neighbor criteria may fail to capture such a kind of significant spatial relationship. In this method, adjustments are made to the neighbor weights based on the spatial distance between neighbors.

First, a spatial weight is defined by using the simple first order neighboring criteria. Then, an average (centroid-to-centroid) distance between the neighboring polygons and polygon i is calculated. A polygon not among those defined as neighbors already will be admitted as a neighboring polygon if its spatial distance (centroid-to-centroid) from the polygon i is less than this average distance.

The element $x[i, j]$ of the weight matrix X is 1 if polygon j is a neighbor of polygon i under the above criteria, and is 0 otherwise.

Higher Order Neighbor Weights

This option constructs a spatial weight matrix of order n based on the spatial neighbor relationship between polygons. Specify the order using the slider bar.

For spatial neighbors of order 1, the element $x[i, j]$ of the resulting weight matrix X is 1 if polygon j is adjacent to polygon i , and is 0 otherwise.

For spatial neighbor weights of order 2, the element $x[i, j]$ of the weight matrix is 1 if polygon j is adjacent to the first order neighbors of polygon i , and is 0 otherwise.

More generally, for spatial neighbor weights of order n , the element $x[i, j]$ of the weight matrix X is 1 if polygon j is adjacent to the neighbors of order $n-1$ of polygon i , and is 0 otherwise. The diagonal elements $x[i, i]$ of the spatial weight matrix X are all equal to 0.

Distance Units

The distance entered in the spatial neighbors dialog should be in terms of the distance units specified for the view; see figure 3.14. The distance units are specified in the ArcView properties dialog. Specify this distance when using one of the methods below for calculating neighbor weights.

Border-to-Border Distance

This option constructs a spatial weight matrix using a border-to-border distance criterion as follows. The element $x[i, j]$ of the weight matrix X is 1 if the shortest distance between the boundary of polygon j and the boundary of polygon i is less than the distance entered, and 0 otherwise.

Centroid-to-Centroid Distance

This option constructs a spatial weight matrix using a centroid-to-centroid distance criterion as follows. The element $x[i, j]$ of the weight matrix X is 1 if the shortest distance between the centroid of polygon j and the centroid of polygon i is less than the distance entered, and 0 otherwise.

In each case, the resulting weight matrix is automatically converted into an S-PLUS object of class `spatial.neighbor`. Specify the name for this object after choosing the method for calculating neighbor weights.

Spatial Autocorrelation

If some aspect of your geographic data is spatially autocorrelated, there may be a need for spatial modeling, such as Spatial Regression. To decide whether spatial modeling should be used, perform a test for spatial autocorrelation. This dialog calculates two common measures of spatial autocorrelation, the Geary and Moran statistics.

Beware of "islands" in your spatial units when computing a spatial neighbor. Records for an "island" must be removed before computation.

The help file for the S+SPATIALSTATS function `spatial.cor` lists the exact formulae for these statistics.

1. Build a list of variables for analysis from fields in the attribute table of the current theme and columns in S-PLUS data frames. You can also type directly in the Variables for Analysis box. The variables must be consistent with each other; they must be associated with the same region of study.
2. Choose a spatial neighbor object in the current S-PLUS workspace. The spatial neighbor object must be consistent with the selected variables; it must be associated with the same spatial region of study. This is usually generated using the dialog described above.
3. Use the slider bar to specify the number of permutations used in the Monte Carlo distribution of the correlation index.
4. Choose between Moran's I or Geary's c correlation indices.
5. Specify a name for the S-PLUS object which is created when the autocorrelation routine is run.
6. A summary report of the routine will be saved in a text file. Specify a name for the report file and indicate whether it should be opened immediately after running the regression. After reviewing the summary information as displayed, be sure to close the opened file so that the dialog will close.



Figure 3.15: *The Spatial Autocorrelation dialog.*

Local Spatial Association

This dialog uses S+SPATIALSTATS to estimate generalized versions of the local Moran and the local Geary indicators of spatial association. The local Moran index helps identification of spatial agglomerative patterns. The local Geary index helps to identify spatial patterns of similarity and dissimilarity. These local statistics are estimated for each row of your data, and can be associated with their global counterparts and used to estimate the contribution of the individual statistics to the corresponding global statistics.

A value is returned for each sampling unit (record); thus the output can be visualized in the View window as another attribute of the corresponding theme.

1. Select one variable for analysis from fields in the attribute table of the current theme and columns in S-PLUS data frames in the S-PLUS data directory. You can also type directly in the Variables for Analysis box.

2. Choose a spatial neighbor object in the current S-PLUS workspace. The spatial neighbor object must be consistent with the selected variable; it must be associated with the same spatial region of study. Use the dialog for spatial neighbors to generate such an object for the current theme.
3. If desired, specify a vector of weights from the list of columns of the currently selected data frame or theme.
4. Choose between Moran's I or Geary's c correlation indices.
5. Specify a name for the S-PLUS object which is created when the correlation routine is run.
6. Check whether the results should be imported as a table into ArcView.
7. Check whether the imported table should be joined to the current theme's attribute table. If so, a join field must be specified.

Caution: If the data does not come from ArcView, the join may not be possible

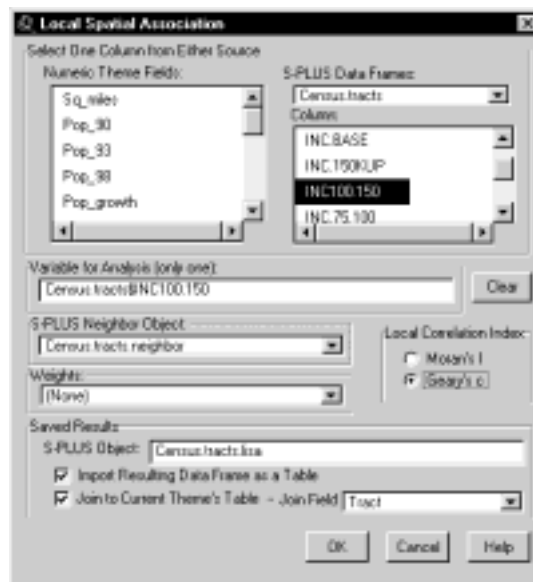


Figure 3.16: The Local Spatial Association dialog.

Spatial Linear Regression

This dialog fits a spatial regression model to selected variables. It looks at both large-scale changes in mean due to spatial location or other explanatory variables, and small-scale variation due to interactions with neighbors.

1. Build a model formula from fields in the attribute table of the current theme and columns in S-PLUS data frames. The variables must be consistent with each other; they must be associated with the same region of study. The first variable selected will be used as the response variable. You cannot type directly into the Model Formula box.
2. Choose a spatial neighbor object in the current S-PLUS workspace. The spatial neighbor object must be consistent with the variables in the model formula; it must be associated with the same spatial region of study. It can be generated using the Spatial Neighbors dialog described earlier.
3. Select one of the following three types of the spatial regression model for the analysis: CAR, or Conditional Autoregressive, SAR, or Simultaneous Autoregressive, and MA, or Moving Average.
4. An S-PLUS data frame object will be saved after the regression runs with fitted and residuals values from the regression. Specify a name for this object and indicate whether it should be imported into ArcView and joined to the active theme. If a join is requested then a Join field must be entered.
5. Indicate whether to include Moran's I test on the estimated residuals.
6. A summary report of the routine will be saved in a text file. Specify a name for the report file and indicate whether it should be opened immediately after running the regression. After reviewing the opened file, close it to complete the operation.

Specific formulae for the covariance matrices corresponding to these models can be found in the help file for the S+SPATIALSTATS function `slm`.



Figure 3.17: The Spatial Regression dialog.

THE HELP SYSTEM

S-PLUS for ArcView Help	While working in ArcView, you can access on-line help with the functionality in the S-PLUS and Spatial Statistics dialogs.
S-PLUS for ArcView User's Manual	Displays an on-line copy of this manual.
S-PLUS Language Reference	This is a comprehensive reference to over 2,000 S-PLUS functions and built-in data sets.
S-PLUS Spatial Statistics Help	Displays help for the S+SPATIALSTATS module
About S-PLUS for ArcView	Provides information about the current version of the extension and the location of the S+SpatialStats module, if installed.

USING S-PLUS AND S+SPATIALSTATS

4

What is S-PLUS	56
Object Browser	56
Graph Sheets	65
Data Windows	69
What is S+SPATIALSTATS	79
Types of ArcView Data and S+SPATIALSTATS	79
Analysis in S+SPATIALSTATS	79
Script Window and Commands Window	81

When you use S-PLUS for ArcView, you have full access to the analytical and visualization tools of S-PLUS, one of the most comprehensive programs available for exploring and analyzing data. The optional S+SPATIALSTATS module extends S-PLUS for modeling spatially correlated data.

With S-PLUS for ArcView, you can use S-PLUS to analyze data from ArcView attribute tables. Tables may be exported to S-PLUS, and analysis results later joined in ArcView to the original theme attribute table. Themes can have features with different geometric shapes. The main shapes that can be analyzed using S+SPATIALSTATS are polygons and points. Analysis is not presently available for themes which represent arcs or other ARC/INFO feature classes.

This chapter illustrates a variety of typical tasks that the S-PLUS for ArcView user will perform in S-PLUS and S+SPATIALSTATS when analyzing GIS data. Starting with data in ArcView, you will perform an analysis in S-PLUS and bring numeric and graphical results back into S-PLUS.

WHAT IS S-PLUS

This section describes several elements of the S-PLUS user interface that you will use when running S-PLUS for ArcView.

- Object Browser
- Data Editor
- Graph sheets
- Command window and Script files

The exercises which follow lead you through the basic operations of S-PLUS. For more detailed and comprehensive information, refer to the *S-PLUS User's Guide*.

Object Browser The S-PLUS environment is object-oriented in that all items are distinct, editable objects. This includes not only data objects and functions, but also graph objects and interface objects, such as menus, dialogs, toolbars, and toolbar buttons. The Object Browser is a simple and powerful interface through which to work with S-PLUS objects. It operates in a way similar to the Windows Explorer. The exercises below illustrate a number of typical uses of the Object Browser in an analysis of data from the Atlanta Census tracts.

Summarizing Census tracts data

If the Object Browser button is already depressed and the Object Browser is not visible, choose Object Browser in the Window menu to bring the Object Browser into view.

1. Open the Object Browser by depressing the Object Browser button . Like the Windows Explorer, it is a two-paned window with collapsible and expandable icons in the left pane.

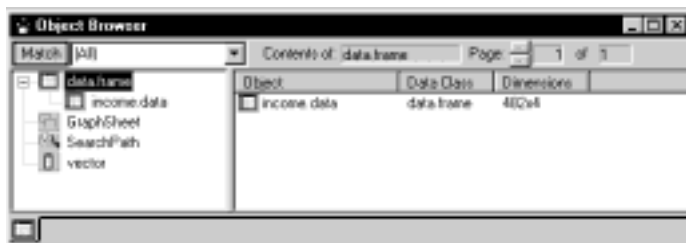


Figure 4.1: *The Object Browser.*

2. Make sure that the right kinds of objects are displayed in the Object Browser. To do this, right-click in the left pane and choose Filtering from the context menu. The Object Browser property dialog appears. Use this dialog to specify what is displayed in the Object Browser.

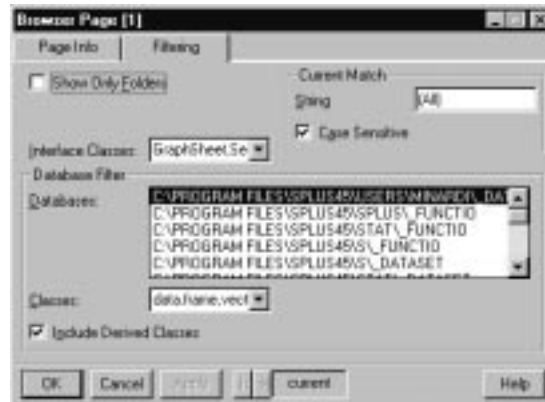


Figure 4.2: *The Object Browser property dialog.*

Next to Databases is a list of directory paths. The top line should be among those selected; this is where S-PLUS for ArcView stores its exported data as S-PLUS objects.

In the list next to Classes, make sure that the data frame item is selected. You may also want to select vectors and functions, say, but avoid choosing too many items in the list, as it may slow down the performance of the Object Browser. The same warning applies to the Databases list. Click OK to apply your choices and close the dialog.

If you chose to display functions, you'll notice the `glisa` function appears under the function icon. This function is currently placed there when S-PLUS for ArcView is installed.

3. Make sure the relevant information about the displayed objects is given. Right-click in the right pane and choose Right Pane from the context menu. Another property dialog appears.

Click the radio button next to List Details. In order to see the names of the selected classes, make sure that the Data Class box is checked. It will also be useful to check the Dimensions box.

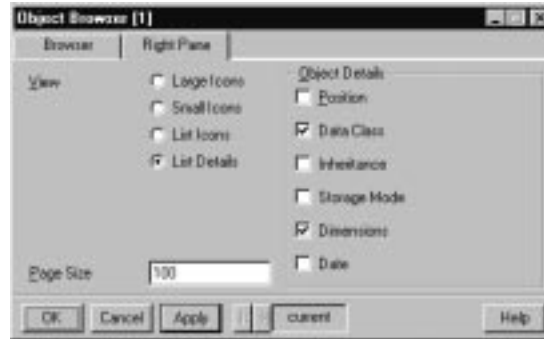


Figure 4.3: The Right Pane property dialog for the Object Browser.

Click OK to apply these selections and close the dialog.

4. *Browse your data.* The Object Browser now shows data which has been exported from ArcView in Exercise 3 of chapter 2. Expand the data frame icon in the left pane and click the data frame *income.data*. Its four fields are displayed, as in figure 4.4. In the Dimensions column, you see that each field has length 482, as expected.

Object	Data Class	Dimensions
Avg inc	double	482
Pct male	double	482
Avg age	double	482
Indicator	integer	482

Figure 4.4: The *income.data* data frame.

Double-click the *income.data* icon to open the table of data in the S-PLUS data editor. To sort on the Indicator field, for example, click anywhere in the Indicator column of data and then click the Sort Ascending button .

5. *Generate summary statistics for the income.data fields.* Click the *income.data* data sheet to make it active. Open the Summary Statistics dialog in the Data Summaries submenu of the Statistics menu. The *income.data* data frame will already be selected, as shown in figure 4.6.



		1	2	3	4	5
		Avg.inc	Pct.male	Avg.age	Indicator	
1	36	-99.00	-99.00	-99.00	0	
2	233	-99.00	-99.00	-99.00	0	
3	236	-99.00	-99.00	-99.00	0	
4	474	-99.00	-99.00	-99.00	0	
5	159	0.00	50.00	38.00	0	
6	32	2500.00	42.86	44.14	1	
7	296	6632.65	88.47	33.15	1	

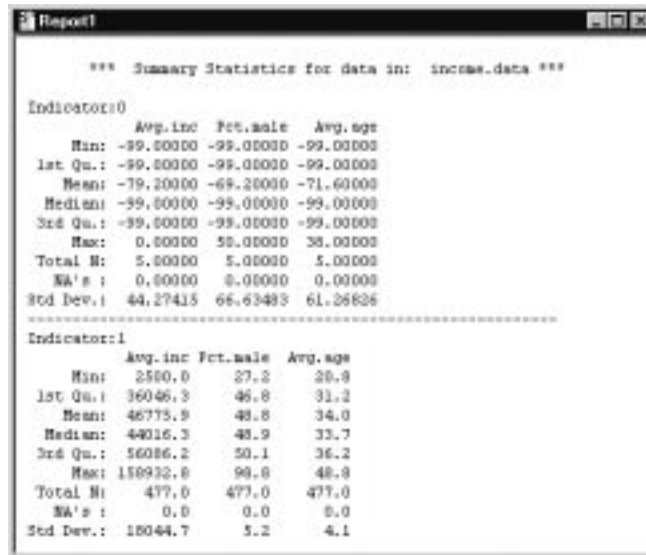
Figure 4.5: *The income.data data frame, sorted by Indicator.*



Figure 4.6: *The Summary Statistics dialog.*

Next to Variables, choose Avg.inc, Pct.male, and Avg.age. Next to Grouping Variables, choose Indicator. Make some choices of statistics to calculate. Be sure that the Print Results box is checked. Click OK.

6. *View the results.* The results are displayed in a Report Window, as seen in figure 4.7. To format output in the report window, just highlight the text, right-click, and choose Font to open the Font



*** Summary Statistics for data in: income.data ***

Indicator:0	Avg.inc	Pct.male	Avg.age
Min:	-99.00000	-99.00000	-99.00000
1st Qu.:	-99.00000	-99.00000	-99.00000
Mean:	-79.20000	-69.20000	-71.60000
Median:	-99.00000	-99.00000	-99.00000
3rd Qu.:	-99.00000	-99.00000	-99.00000
Max:	0.00000	50.00000	38.00000
Total N:	5.00000	5.00000	5.00000
NA's :	0.00000	0.00000	0.00000
Std Dev.:	44.27415	66.63483	61.26826

Indicator:1	Avg.inc	Pct.male	Avg.age
Min:	2500.0	27.2	20.9
1st Qu.:	36046.3	46.8	31.2
Mean:	46775.9	48.8	34.0
Median:	44016.3	48.9	33.7
3rd Qu.:	56086.2	50.1	36.2
Max:	150932.0	90.8	48.9
Total N:	477.0	477.0	477.0
NA's :	0.0	0.0	0.0
Std Dev.:	18044.7	5.2	4.1

Figure 4.7: The output from the Summary Statistics dialog.

dialog. Highlighted text in the Report Window can be copied to the clipboard and pasted into another program. You can even paste the text into an ArcView layout.

7. *Trim the data using the Indicator field.* When analyzing the data below, you will want to exclude the records which we have marked with 0 in the Indicator field. Click again on the income.data icon in the Object Browser and choose Subset from the Data menu. The dialog should show income.data next to Data Frame. Select Avg.inc, Pct.male, and Avg.age next to Columns in Subset. To include those records for which Indicator is 1, enter Indicator = 1 next to Subset Rows with. Select Data Frame as the Result type, and enter income.trim as the name of the new data frame next to Save As. Click OK.

The Object Browser shows that the new data frame, income.trim, has 477 records and 3 fields, as expected.

Running Linear Regression

1. *Find the data fields.* Expand the data frame icon in the left pane of the Object Browser to show income.trim. In the right pane, the three fields of income.trim are listed.



Figure 4.8: *The Subset dialog.*

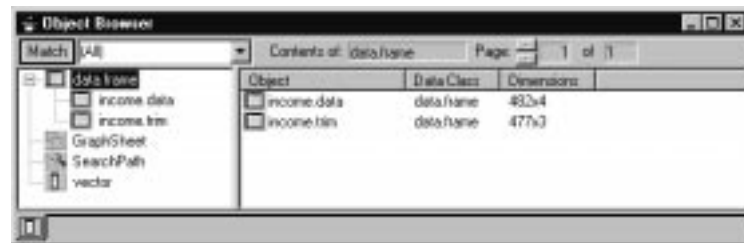


Figure 4.9: *The Object Browser showing income.data and income.trim.*

2. *Specify a Linear Regression model.* Ctrl-click the fields Avg.inc, Pct.male, and Avg.age, in that order.

3. Open the Linear Regression dialog, seen in figure 4.10. It is in

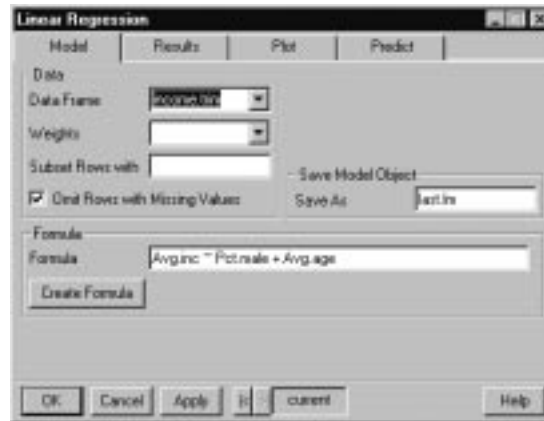


Figure 4.10: The Linear Regression dialog.

- the Regression submenu of the Statistics menu. Next to Formula, you see the formula which you have specified. The first selected column, Avg.inc, is the dependent variable. Pct.male and Avg.age are the independent variables. You may click the button Create Formula to build the formula using the variables in the data frame income.trim.
4. *Save the results of the fit.* Move to the Results page. Check the box next to Long Output under Printed Results. Check the box next to Residuals and enter income.res next to Save In under Saved Results.
5. *Specify plots to be generated.* Move to the Plot page. Check the boxes next to Residuals vs Fit and Residuals Normal QQ. These diagnostic plots help to determine whether the results of the linear regression on your particular data can be trusted.
6. *Click OK* to run the regression, print the results, save the fit, and generate plots.

Using the results of Linear Regression

1. *View the results of the linear regression.* After running the linear regression, you will see a Report Window with printed results and a two-page Graph Sheet, as shown in figure 4.11. From the output in the Report Window, it seems that both Pct.male and Avg.age are related to Avg.inc in a linear fashion. However, the

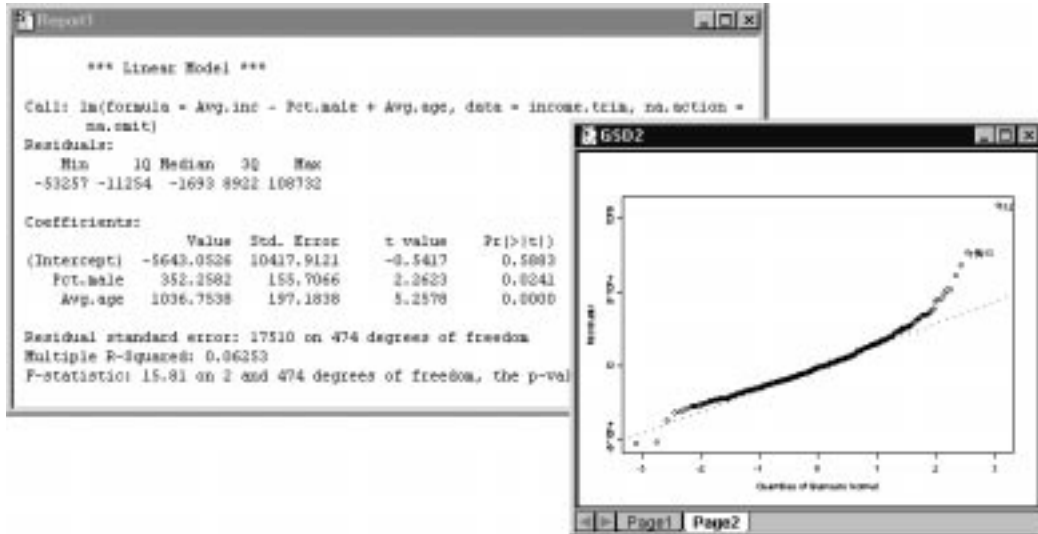


Figure 4.11: Results of the linear regression.

normal quantile-quantile plot indicates that the residuals do not follow a normal distribution. (They have a long tail on the high side.) Based on this, the results of this linear regression are suspect.

2. *Make a folder to group data of interest.* Click the Object Browser window to make it active. Right-click in the left pane and choose Insert Folder from the context menu. Type Census data for the name of the folder. Click the income.trim icon in the left pane, then drag Avg.age and Pct.male into the Census data folder. Click the income.res icon in the left pane and drag residuals into the Census data folder.
3. *Plot the residuals against Avg.age.* Ctrl-click Avg.age, and then residuals, in the Census data folder. Depress the 2D Plots button , if it is not already depressed, to open the 2D Plot Palette. In the 2D Plot Palette, click the Scatterplot button . A Graph sheet opens, showing a scatterplot of residuals against Avg.age, as seen in figure 4.12. From this plot it is seen that the residuals increase with increasing values of Avg.age. When linear regression is valid, the scatter of the residuals is relatively homogeneous along each independent variable.

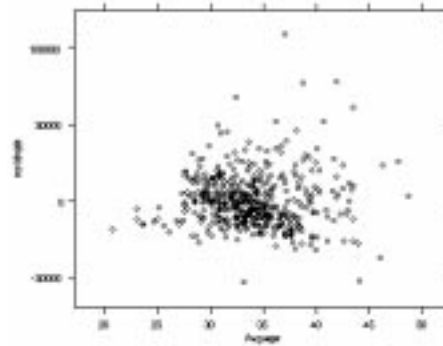


Figure 4.12: Plot of residuals against Avg.age.

4. *Transform Avg.inc.* To remedy the problem seen above, you can try replacing the Avg.inc by a transformation of it which grows more slowly. A common choice is to take the logarithm of the data. Click income.trim in the left pane of the Object Browser and choose Transform from the Data menu. The Transform dialog appears.

Enter Log.avg.inc as the new column name. Notice Avg.inc in the Variable box. Select log10 in the list next to Function and click Add. The S-PLUS expression is automatically filled in, as seen in figure 4.13. Click OK.



Figure 4.13: The Transform dialog.

5. *Perform another linear regression.* Click Log.avg.inc, then Ctrl-click Avg.age and Pct.male in the right pane of the Object Browser. Open the Linear Regression dialog.
6. *Save the results of the fit.* Move to the Results page. Check the box next to Long Output. Check the box next to Residuals and

enter log.income.res next to Save In under Saved Results.

7. *Specify plots to be generated.* Move to the Plot page. Check the boxes next to Residuals vs Fit and Residuals Normal QQ.
8. *Click OK* to run the regression, print the results, save the fit, and generate plots. The plots of Residuals vs Fit and the Residuals Normal QQ are much improved, as shown in figure 4.14. The results of the linear regression are much more reliable. Notice from the Report Window output, however, that the linear regression only explains about 4.5% of the variation in Avg.inc.

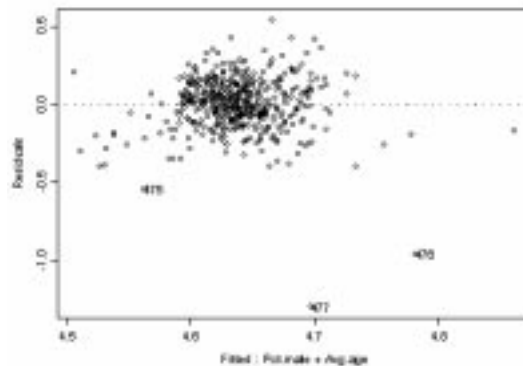


Figure 4.14: *Plot of Residuals against Fitted Values.*

Residual standard error: 0.1693 on 474 degrees of freedom
 Multiple R-Squared: 0.04543
 F-statistic: 11.28 on 2 and 474 degrees of freedom,
 the p-value is 0.00001639

Graph Sheets

In S-PLUS, plots are placed in Graph sheets, just as plots and charts are placed in ArcView layouts. The general technique for editing a plot in an S-PLUS Graph sheet is to right-click on the region of interest, open a property dialog, and make the desired change to the plot. The following exercise illustrates several typical ways to customize a plot and its Graph sheet. For more details, refer to the *S-PLUS User's Guide*.

Diagnosing the Residuals

1. *Select the data.* Click the income.res icon in the left pane of the Object Browser, move to the right pane, and click the residuals field.

2. *Open the 2D Plots palette.* Depress the 2D Plots toolbar button . The 2D Plots palette opens. (It may be open already, if you followed the last exercise.)
3. *Generate the histogram.* Click the Histogram button  in the 2D Plots palette. The histogram is generated, as shown in figure 4.15.

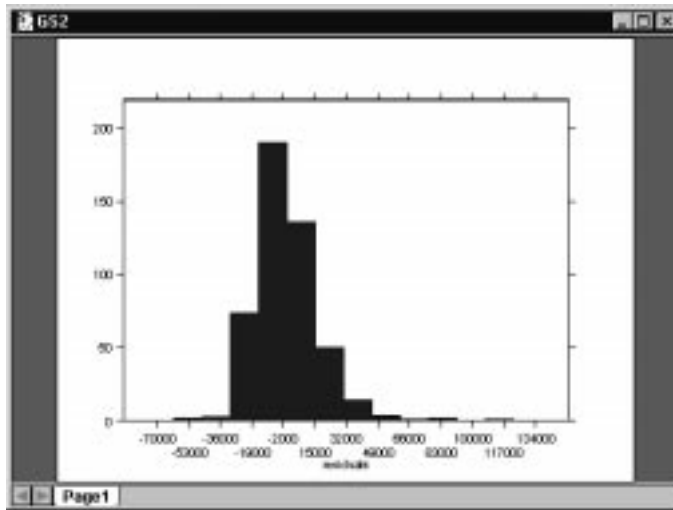


Figure 4.15: *Histogram of residuals.*

4. *Customize the bars.* It would be nice to see the divisions between the bars more clearly. Right-click on the histogram bars. You have clicked in the right spot if a green handle appears below the leftmost bar. Choose Histogram Bars from the context menu. Check the box next to Draw Bars and click OK. Now borders are drawn around the histogram bars.
5. *Change the number of bars.* To better see the features of the histogram, regenerate it with more classes. Right-click the histogram again and choose Options from the context menu. Next to Number of Bars, enter 24.

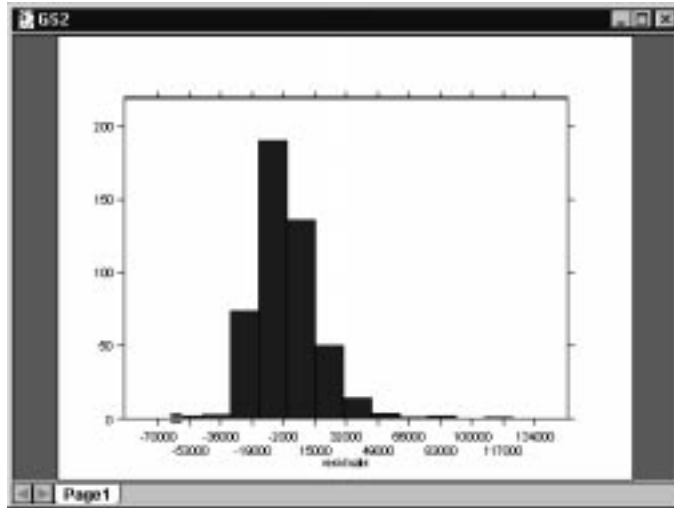


Figure 4.16: Histogram with borders around bars.

6. *Create a title.* Place a title on the histogram by choosing Main from the Title submenu of the Insert menu. A text box appears in the Graph sheet. Type Histogram of Residuals and click in another area of the Graph sheet. The result is shown in figure 4.17.

Visualizing a Linear fit in 3 Dimensions

1. *Select the model.* Click the income.trim icon in the left pane of the S-PLUS Object Browser, move to the right pane, and ctrl-click the fields Pct.male, Avg.age, and Log.avg.inc, in this order.
2. *Open the 3D Plots palette.* Depress the 3D Plots toolbar button . The 3D Plots palette opens.
3. *Generate the plot of the linear fit.* Click the 3D Regression Scatter button . A three dimensional scatter plot is generated, along with a gridded plane representing the Linear Regression fit.
4. *Rotate the plot.* To view the fit from another angle, rotate the three dimensional plot. Click any white space within the plotting area. Five handles, four round and one triangular, will appear. To rotate about the vertical axis, drag one of the round

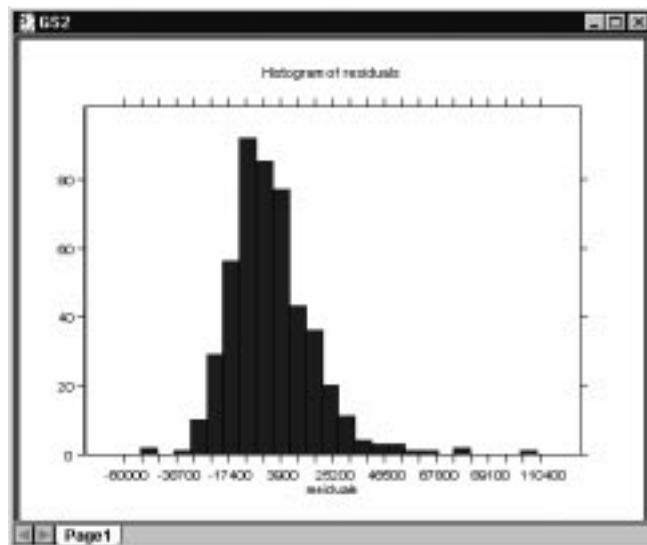


Figure 4.17: Histogram with 24 bars and a title.

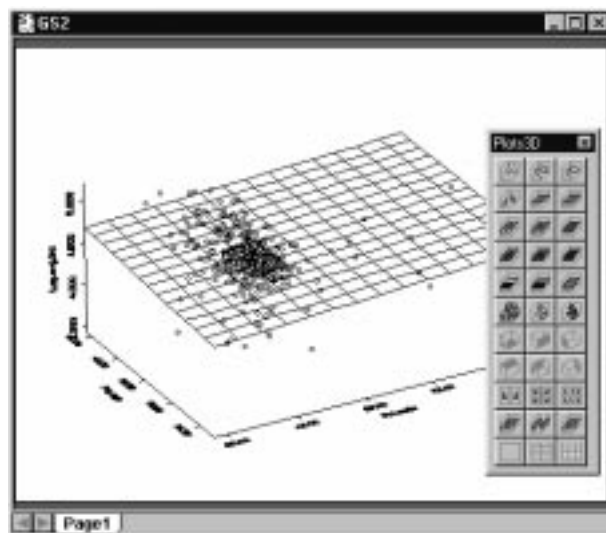
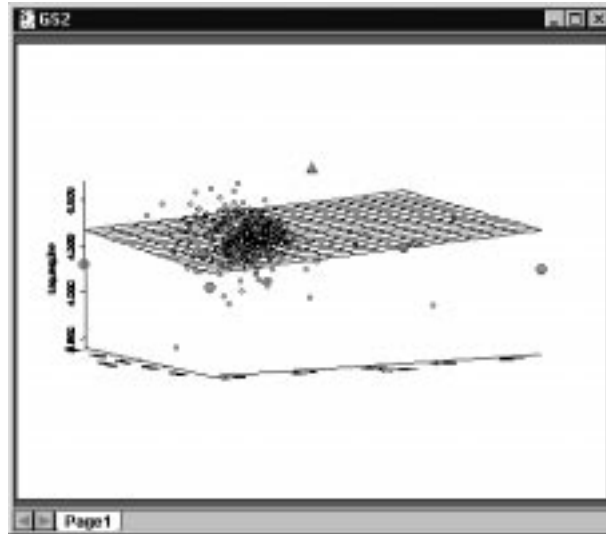


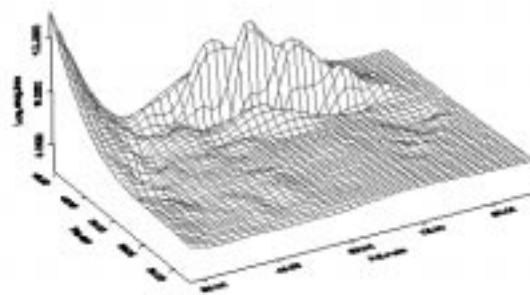
Figure 4.18: Graph sheet showing the three dimensional fit and the 3D Plots palette.

handles. To rotate into the graph sheet, drag the triangular handle. A different view, with handles visible, is shown in

figure 4.19.

**Figure 4.19:** A rotated view of the three dimensional fit.

5. *Create a spline plot.* To see a nonlinear fit of the same data, click the Surface Spline button  in the 3D Plots palette. The spline plot brings out features which are not evident from the three dimensional scatterplot. This is shown in figure 4.20.

**Figure 4.20:** Spline plot showing a nonlinear fit of *Log.avg.inc* by *Avg.age* and *Pct.male*.

Data Windows

An S-PLUS Data window is similar to a spreadsheet, but is column-oriented rather than cell-oriented. You will use Data windows in S-PLUS to

browse, edit, and transform data. The following exercise illustrates a number of common tasks with Data windows. In the exercise you will also see how to perform an analysis of variance in S-PLUS and bring the results back into ArcView.

Export Data to S-PLUS

1. In ArcView, open the *usa.apr* project. (You can find the *usa.apr* project in the **/esridata** directory.) Open the Continental United States view and click the U. S. States theme in the table of contents.
2. *Export the attribute table.* Choose Export Data to S-PLUS from the S-PLUS menu and export the theme table to an S-PLUS data frame called US.States.
3. *Select State_name as the Join Field.* From the dropdown box with field names, select one to be used for future joins, say

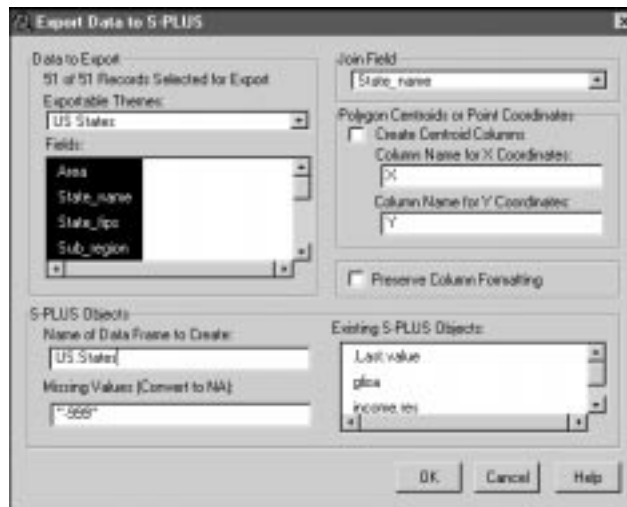


Figure 4.21: Exporting the US States theme table to S-PLUS.

State_name. Click OK. The US.States data frame is now visible in the Object Browser.

The data frame state now appears in a Data window. Figure 4.23 shows the imported data frame.

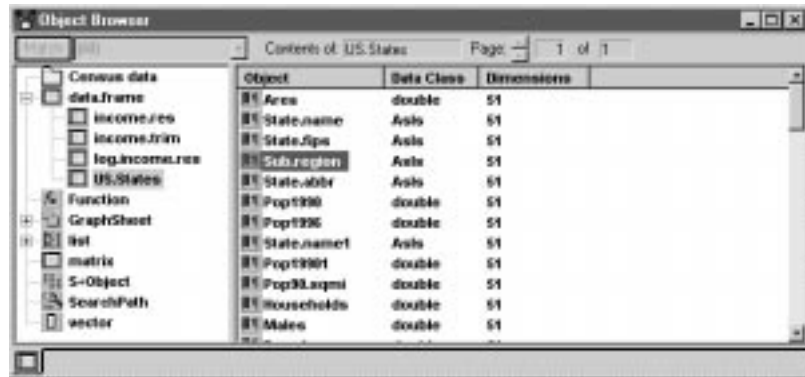


Figure 4.22: The newly imported *US.States* data frame.

Perform an Analysis of Variance

Analysis of variance is used to determine whether the average values of a continuous field, like average income, vary between categories of a discrete field, like state. Here we analyze the effect of geographical sub-region on the median rent in a state.

4. *Build a model.* Begin by highlighting the fields of interest in the data frame *US.States*. First click the field *Medianrent*, then Ctrl-click the field *Sub.region*.

	39	40	41	42	43	44
	Owner.occ	Renter.occ	Median.val	Medianrent	Units.1det	Units.1
1	1171580.00	700851.00	93400.00	383.00	272721.00	48081
2	205899.00	100264.00	56600.00	251.00	237533.00	8431
3	327888.00	137424.00	87400.00	358.00	378413.00	11751
4	157950.00	82928.00	50800.00	266.00	172938.00	10281
5	171161.00	87873.00	45200.00	242.00	202166.00	5241
6	114544.00	54295.00	61600.00	270.00	129197.00	6211
7	1215350.00	808789.00	81500.00	323.00	342130.00	40331

Figure 4.23: Highlighting columns in *US.States*.

5. *View the model formula.* Choose the Fixed Effects Analysis of Variance dialog from the Statistics menu. The model page is showing when it appears. The *US.States* data frame is already chosen, and the formula field reflects the currently highlighted columns; the first chosen field, *Medianrent*, is taken as the

response variable and is placed on the left side of the formula, and the explanatory factor Sub.region is placed on the right side (see figure 4.24).

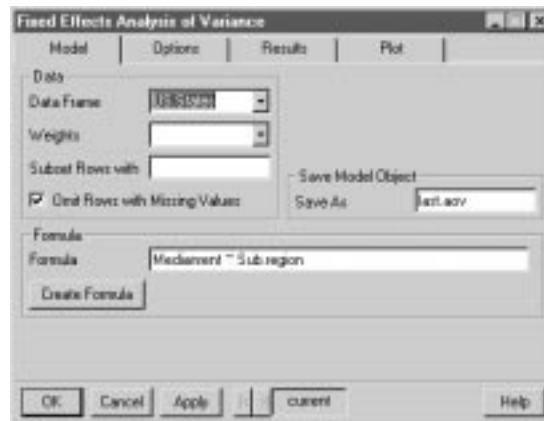


Figure 4.24: The Model page of the Fixed Effects Analysis of Variance dialog.

6. *Save relevant results.* Click the Results tab of the dialog. Select the printed results Short Output and Type I Sums of Squares. Save the resulting fitted values and residuals in a data frame called state.anova.res (see figure 4.25).

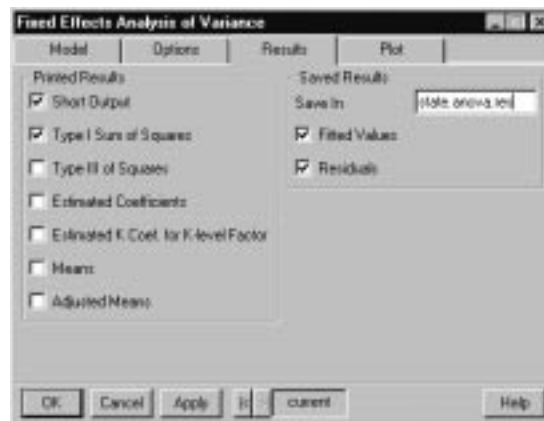


Figure 4.25: Results page of the Fixed Effects Analysis of Variance dialog.

7. *Create diagnostic plots.* Click the Plot tab of the dialog. Specify the Residuals vs Fit and Residuals Normal QQ plots. Click OK.

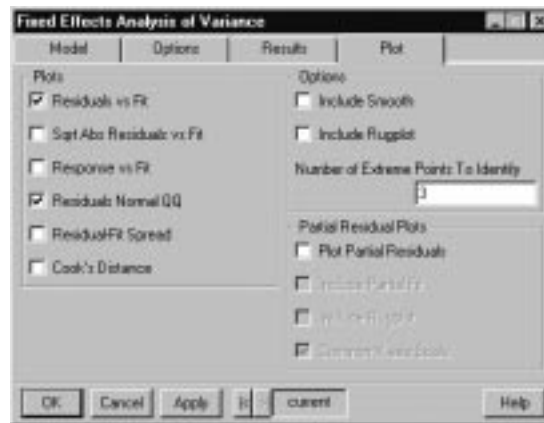


Figure 4.26: Plot page of the Fixed Effects Analysis of Variance dialog.

8. *View the plotted results.* The plots appear on pages of a Graph sheet. The plot of fitted values against residuals shows a relatively homogeneous scatter of residuals across fitted values. There are a few slightly outlying values; the three most extreme are labeled by row number.

The plot of the residuals against normal quantiles does not indicate any marked departure from the assumption that the residuals are normally distributed. The plots suggest that the results of the Analysis of Variance can be trusted.

9. *View the printed results.* The printed output appears in a Report window. The F statistic of 6.645 on 8 and 42 degrees of freedom yields a p-value of .000014; this indicates a significant difference in the mean values of Medianrent between subregions.

Now you will prepare the results of the Analysis of Variance for import into ArcView. Keep in mind that the rows of the US.States data and state.anova.res data frames correspond.

10. *Work with the saved results.* Open the data frames state.anova.res and US.States. Highlight the column State.name in US.States, click in the rows of data, and drag over an empty column in state.anova.res, as shown in figure 4.29. This copies that column into the state.anova.res data frame. Copy the

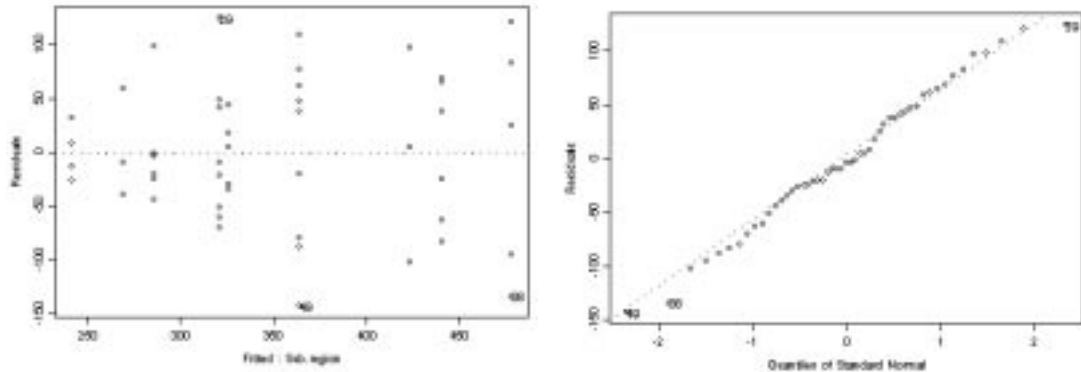


Figure 4.27: Graphical output of the Analysis of Variance.

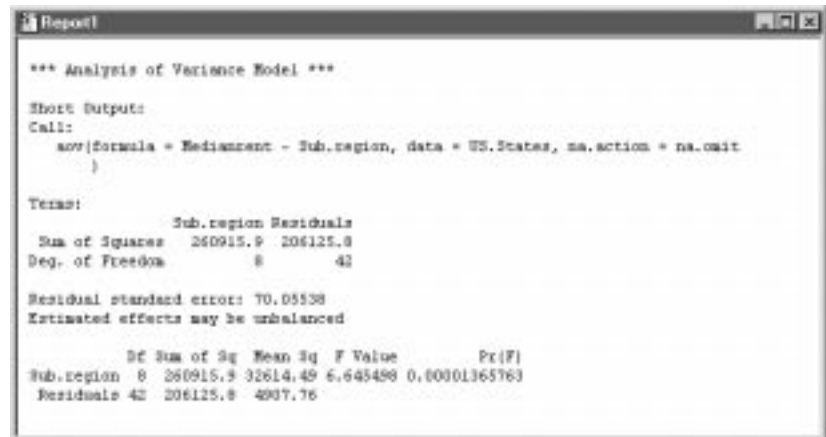


Figure 4.28: Printed output of the Analysis of Variance.

column Sub.region from US.States into state.anova.res as well.

11. Are the fitted values reasonable? Highlight the Sub.region column in state.anova.res and click the Sort Ascending toolbar button . Scroll down the data frame and notice that the fitted values are the same within each subregion, as expected (see figure 4.30).

Highlight the Sub.region column once more, and click the Remove Column toolbar button  to remove the column from state.anova.res.

Figure 4.29 shows two S-PLUS windows. The top window, titled 'state.anova.res', displays a table with two columns: 'fit' and 'residuals'. The bottom window, titled 'US.States', displays a table with six columns: 'Area', 'State.name', 'State.fips', 'Sub.region', 'State.abbr', and 'Pop.'. The 'State.name' column in the 'US.States' window is highlighted, indicating it is being dragged to the 'state.anova.res' window.

Figure 4.29: Using drag and drop to copy the *State.name* column from *US.States* into *state.anova.res*.

Figure 4.30 shows the 'state.anova.res' window after adding the 'State.name' column. The table now has four columns: 'fit', 'residuals', 'State.name', and 'Sub.region'. The 'Sub.region' column is highlighted.

Figure 4.30: Using the additional column *Sub.region* to verify results from Analysis of Variance.

Bring the results into ArcView for visualization

12. Move to ArcView now and choose Import Table from S-PLUS in the S-PLUS menu. Import the *state.anova.res* data frame. Specify Join Imported Columns to Theme and select US States as the theme to join from the dropdown box.

13. *Match records based on common values in ...*Select State_name as the Imported Column, and the same as the Theme's Field. Click OK (see figure 4.31).

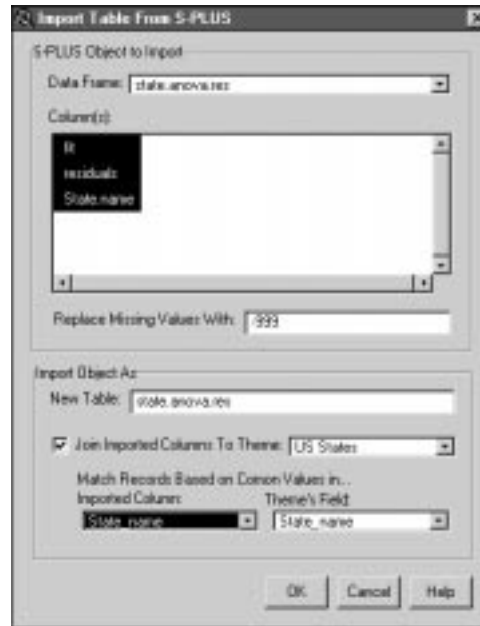


Figure 4.31: Using the Import Table from S-PLUS dialog in ArcView to bring the results from the Analysis of Variance into ArcView.

14. *View the results.* To view the results of the Analysis of Variance in the Continental United States view, you may use the Color Classification dialog from the S-PLUS menu in ArcView. Select US States as the active theme and Residuals (last column) as the field or column to classify upon. Click OK (see figure 4.32).

To change color scale, you may double-click the US States theme name in the table of contents of the view. The states in the view are lightly shaded where median rent is relatively low for the subregion, and heavily shaded where median rent is relatively high for the subregion (see figure 4.33).

One feature seems unusual in this view; the cluster of very lightly shaded states covering the Pacific Northwest. In a given region, one would expect some states with high median rent,

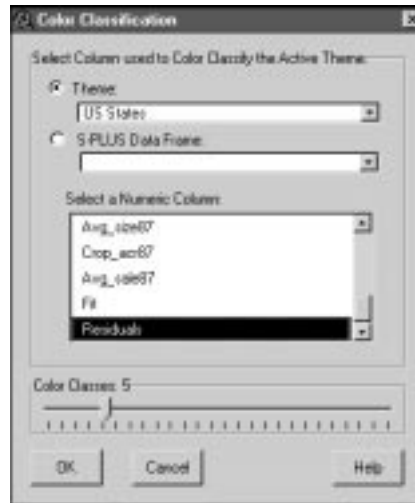


Figure 4.32: Using the Color Classification dialog to visualize the results from the Analysis of Variance performed in S-PLUS.

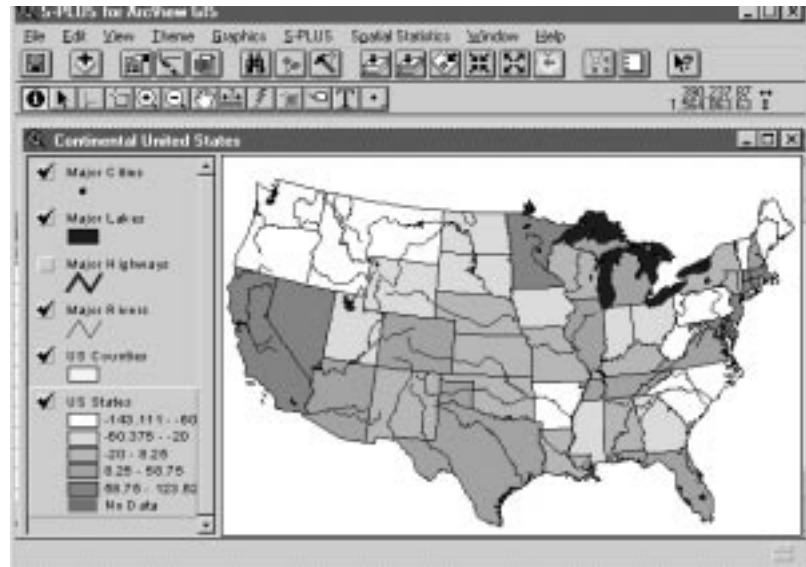


Figure 4.33: A view of the U. S. States with a graduated color legend type. States are classified by Residuals, taken from the Analysis of Variance in S-PLUS.

and some states with low median rent, relative to the subregion as a whole. A look at the attribute table reveals that the "Pacific" subregion which contains Washington State also contains Oregon, California, Alaska, and Hawaii. Choose Full Extent from the Zoom menu. Now Alaska and Hawaii are in view; they appear to have relatively high median rent values compared to their subregional neighbors, as expected (see figure 4.34).



Figure 4.34: An expanded view of the U. S. States with a graduated color legend type.

WHAT IS S+SPATIALSTATS

Spatial structure is intrinsic to ArcView data. With spatially-oriented data, the assumptions of independent observations used in statistical inference are typically not valid. In exercise 3 of chapter 2, for example, we saw demographic data from the Atlanta census tracts that are highly correlated in space. In spatial analysis, this spatial structure is incorporated into the statistical model. The S+SPATIALSTATS module extends the functionality of S-PLUS with such tools for modeling data having spatial dependencies.

Types of ArcView Data and S+SPATIALSTATS

S-PLUS for ArcView works with spatial data stored in ArcView. Themes can have features with different geometric shapes; the main shapes that can be analyzed using S+SPATIALSTATS are polygons and points. Analysis is not presently available for themes which represent arcs or other ARC/INFO feature classes.

Points can be exported to S-PLUS for ArcView and analyzed in S+SPATIALSTATS as point patterns, if you are interested in detecting specific patterns in the locations themselves. With S+SPATIALSTATS tools you typically explore whether the points are randomly distributed in space or whether they follow specific patterns.

Polygons are the focus of the Spatial Statistics menu in S-PLUS for ArcView. Attributes of polygonal features are called lattice data in S+SPATIALSTATS. The neighbor structure among these features is integrated into the models in the S+SPATIALSTATS module, and in the Spatial Statistics menu items.

Analysis in S+SPATIALSTATS

The S+SPATIALSTATS functions can be separated along the lines of the three broad categories of spatial data on which they act.

- Geostatistical data, or random field data
- Lattice data
- Spatial point patterns

The assignment is not meant to be exclusive, since many of the functions can be used on different types of spatial data. For example, spatial regression modeling is typically used for lattice data, but it can be used on geostatistical data. Likewise, it is possible to perform geostatistical-type analyses, such as variogram estimation, on lattice data.

In ArcView, point themes represent geostatistical and point pattern data, while polygons represent lattice data.

For analyzing **geostatistical** data, S+SPATIALSTATS includes tools to perform the following:

- Estimate and visualize standard or robust, omnidirectional or directional variograms
- Model empirical variograms; fit theoretical variogram models to empirical data
- Perform ordinary kriging to obtain point estimates for unsampled locations and kriging prediction variances
- Perform universal kriging to model large-scale trends while calculating predictions.

For the analysis of **lattice** data, S+SPATIALSTATS includes functions to perform the following:

- Find nearest neighbors or groups of neighbors based on distance or common boundaries
- Calculate and test for spatial autocorrelation using the Moran and Geary correlation coefficients
- Perform spatial regression modeling using conditional autoregressive, simultaneous autoregressive, or moving-average covariance structures.

For the analysis of **point pattern** data, S+SPATIALSTATS includes functions to perform the following:

- Plot geometrically accurate point maps without spatial distortion
- Explore complete spatial randomness using nearest neighbor methods
- Estimate intensity using kernel, loess, and gaussian methods
- Test second-order stationarity using Ripley's K-function.

S+SPATIALSTATS also includes methods for simulating the three basic types of spatial data. For a detailed example, see chapter 5.

Script Window and Commands Window

Functions in the S+SPATIALSTATS module are currently available only at the command line in S-PLUS or in the Spatial Statistics menu in S-PLUS for ArcView. This section illustrates how to analyze your ArcView data with S+SPATIALSTATS from the command line in S-PLUS. Turn to chapter 5 to see how to use the Spatial Statistics menu in S-PLUS for ArcView. Future releases of S+SPATIALSTATS will have full functionality from the S-PLUS user interface.

In this section, you will learn how the S+SPATIALSTATS module can be used for the analysis of your ArcView data. This includes:

- how to use the Commands and Script windows in S-PLUS for the analysis of spatial data using S+SPATIALSTATS
- what types of analysis are available in S+SPATIALSTATS.

The Commands and Script windows in S-PLUS give you access to the full power of over 2,000 S-PLUS functions. Use the Commands window to quickly run commands singly. Use the Script window to run and edit several commands at a time, and to save sets of commands as scripts for future use or distribution to others. Here are a few simple examples using the SIDS data.

If the Commands Window button is depressed and the Commands Window is not visible, choose Commands in the Window menu to bring it forward.

1. Open the Commands window by depressing the Commands Window toolbar button .
2. If it is installed, the S+SPATIALSTATS module is automatically loaded when S-PLUS for ArcView starts up. When you run S-PLUS without running ArcView, you must load S+SPATIALSTATS module. Do this by entering the following command in the Commands window.

```
> module(spatial)
```

You can check that the S+SPATIALSTATS module is loaded by entering

> search()

and checking that the S+SPATIALSTATS data directory is in the second position in the list of S-PLUS databases, as seen in figure 4.35. To rerun a previous command, press the Up arrow until it reappears.



Figure 4.35: The Commands window.

3. To open a Script window to manage a short analysis, click the New toolbar button . Select Script File from the New dialog and click OK. A new Script window opens.
4. In chapter 5, you work with data on SIDS rates in 100 counties in North Carolina. The first two commands in the top pane of the Script window in figure 4.36 create the SIDS data frame in S-PLUS and identify the location of the built in spatial neighbor object `sids.neighbor`. The next two commands calculate an estimate of Spatial Correlation. You may highlight those commands and click the Run button or press F10 to run them. The output appears in the bottom pane of the Script window, as seen in figure 4.36.
5. Click the Run button again to rerun the lines. Some of the output is not identical because random permutations are performed in generating the sample autocorrelation. Similarly, your quantiles of permutations-correlations and the corresponding p-value may not match that in figure 4.36. Both the Moran statistic and the permutation test give evidence



```

source("c:\\esri\\arcgis\\arcview\\samples\\arcview\\sids.q")
find(sids.neighbor)
sids.cor <- spatial.cor(sids$sid.ft * sqrt(sids$births),
  neighbor=sids.neighbor, statistic="moran",
  sampling="free", npermut=1000)
sids.cor

> sids.cor <- spatial.cor[sids$sid.ft * sqrt(sids$births), neighbor =
  sids.neighbor, statistic = "moran", sampling = "free",
  npermut = 1000]
> sids.cor

      Spatial Correlation Estimate

Statistic = "moran" Sampling = "free"

Correlation = 0.2287
Variance    = 0.03772
Std. Error  = 0.08796

Normal statistic = 2.717
Normal p-value (2-sided) = 0.005579

Null Hypothesis: No spatial autocorrelation

Summary of the permutation-correlations :
      Min. 1st Qu. Median Mean 3rd Qu. Max.
-0.4026 -0.06709 -0.01518 -0.01254 0.03986 0.6568

permutation p-value = 0.811

```

Figure 4.36: Running two commands from the Script window.

against the null hypothesis of no spatial autocorrelation. For more discussion of this, see section 5.2 of the *S+SPATIALSTATS User's manual*.

6. Click the Save toolbar button  to save the script in a file for later use or distribution.

ANALYZING SPATIAL DATA IN S-PLUS FOR ARCVIEW

5

Spatial Statistics	86
Data Preparation	86
Defining Neighbor Relationships	89
Spatial Autocorrelation	91
Local Spatial Association	94
Spatial Regression	98
Bibliography	103

This chapter provides a detailed example of spatial data analysis in S-PLUS. In several stages you will see the unique way in which S-PLUS and ArcView GIS work together. If you are unfamiliar with S-PLUS, look at chapter 4, Using S-PLUS and S+SPATIALSTATS, before continuing here. Chapter 4 describes the basics of effective data management and analysis in S-PLUS.

Data coming from a GIS package has an inherent spatial component that is of central interest to the data analyst. In this chapter, spatially oriented data is used in an example which requires the optional S+SPATIALSTATS module to S-PLUS. Using these sample data, you will see how to:

- perform exploratory data analysis using the Commands window
- define neighbor relationships for use in spatial analysis
- calculate spatial autocorrelation and local association measures
- perform spatial regression

SPATIAL STATISTICS

This exercise uses data provided with the S+SPATIALSTATS module, complemented with data from ArcView. The data frame `sids` contains rates of death from Sudden Infant Death Syndrome (SIDS) for the years 1974-1978 (Cressie and Chan, 1989) for each of the one hundred counties in the state of North Carolina.

Exploratory analysis of this data can be found in Chapter 3 of the *S+SPATIALSTATS User's Manual*. Chapter 5 of that manual describes several approaches to the analysis of these data. In this section, we will use the same data to illustrate the spatial analysis tools available in S-PLUS for ArcView.

Data Preparation

If the Commands Window button is depressed and the Commands Window is not visible, choose Commands in the Window menu to bring it forward.

Caution: Some typographical errors existed in the original data distributed with S+SPATIALSTATS Version 1.1. Use the data set described here for reliable results.

1. *Open the Commands window in S-PLUS.* Open the Commands window by depressing the Commands Window toolbar button . For information on the Commands window, refer to the Script Window and Commands Window section in chapter 4. In what follows, commands are indicated next to the default S-PLUS prompt `>`. Parts of commands which exceed the length of a line are preceded by the default S-PLUS continuation prompt `+`. When working along with the examples, type only the command, and not the prompts.
2. *Load the S+SPATIALSTATS module, if available.* This is not needed in all the examples in this section. If it is installed, the S+SPATIALSTATS module is automatically loaded when S-PLUS for ArcView starts up. When you run S-PLUS without running ArcView, you must load S+SPATIALSTATS module. Do this by entering the following command in the Commands window.

```
> module(spatial)
```

You can check that the S+SPATIALSTATS module is loaded by entering

```
> search()
```

and checking that the S+SPATIALSTATS data directory is in the second position in the list of S-PLUS databases, as seen in figure 5.1.



Figure 5.1: The Commands window.

3. *Bring the data into S-PLUS.* To bring the data into S-PLUS, open the file **\$AVHOME\Samples\SarcView\sids.q** in S-PLUS, where **\$AVHOME** refers to the directory in which ArcView is installed. This is typically **c:\esri\av_gis30\arcview**, as we assume here.

Enter the following command in the Commands window and hit Enter:

```
> source("c:\\esri\\av_gis30\\arcview\\samples\\sarcview\\sids.q")
```

This adds the data frame `sids` to your S-PLUS working directory. You can check that you have now two versions of the `sids` data frame by running

```
> find(sids)
[1] "C:\\Program Files\\splus45\\users\\minardi\\_Data"
[2] "C:\\Program Files\\splus45\\module\\spatial\\_Data"
```

The output indicates that there are two copies of the `sids` data frame. The first one listed, in the default S-PLUS working directory, was loaded by the previous command. It will be used in our analysis.

Remove or rename the first version (the one that was just sourced in) when you want to follow the examples in the *S+SPATIALSTATS User's Manual*, otherwise you may encounter errors as you go due to the discrepancies in county names, and the precedence of this first version over the one already in S+SPATIALSTATS.

The variables in the SIDS data frame are displayed as follows.

```
> names(sids)
[1] "id"          "easting"      "northing"     "sid"
[5] "births"      "nwbirths"     "group"        "sid.ft"
[9] "nwbirths.ft"
```

For more detail on each variable, look at the help file for sids.

```
> help(sids)
```

4. *Create an ArcView shapefile with county information.* Move to ArcView. Open the **usa.apr** project. You may find this in the **c:\esri\esridata** directory parallel to your \$AVHOME directory. Bring up the table of Attributes for the US Counties theme. Using the Query Builder toolbar button (the "hammer" tool), select all counties such that the field State_name equals "North Carolina". Click on the view with the US Counties theme and from the Theme menu select Convert to Shapefile... to create a new shapefile with the data for North Carolina only. Call the new file **ncco.shp**.

This new shapefile will be added to the view's table of contents. Make it active by clicking on it and check-marking it. You may zoom the view in to this active theme.

Export the County names into S-Plus

5. *Choose Export Data to S-PLUS from the ArcView menu.* Select Name below Exportable Fields. Enter Ncco.names under Name of Data Frame to Create. Click OK.
6. *Move to S-PLUS.* Ncco.names appears in S-PLUS as a data frame with one factor column. Turn this factor column into a character vector of the county names with the following command. Notice that S-PLUS object names are case-sensitive.

```
> ncco.names <-
+   as.character(Ncco.names[,1])
```

The county names for the sids data frame are the row names of the data frame. You can get those as follows:

```
> sids.names <- row.names(sids)
```

A quick test shows that the counties are sorted differently in each vector, even though they both contain the same strings.

```
> all(sids.names == ncco.names)
[1] F
> all(sort(sids.names)==sort(ncco.names))
[1] T
```

7. *Reorder SIDS data frame accordingly.* The rows of the `sids` and `ncco.names` data frames must be joined on the names of the counties. To this end, you want to find out the order of the names in `ncco.names` and reorder the `sids` data frame accordingly. The S-PLUS function `match` returns a vector of the positions of `sids.names` in `ncco.names`. Once we have these positions, re-ordering `sids` as `sids.arc` will yield the desired result, a data frame with the same ordering as `ncco.names`, hence the same ordering of the records in ArcView.

You should compare these 10 values to the first 10 rows of the corresponding table in ArcView.

```
> ind <- match(ncco.names,sids.names)
> sids.names[ind][1:10]
[1] "Ashe"           "Alleghany"      "Surry"
[4] "Currituck"     "Northampton"   "Hertford"
[7] "Camden"         "Gates"          "Warren"
[10] "Stokes"
> all(sids.names[ind]==ncco.names)
[1] T
> sids.arc <- sids[ind,]
> all(row.names(sids.arc) == ncco.names)
[1] T
```

Now we can proceed with our spatial analysis. Note that this data frame can also be imported into ArcView as a point theme using its columns easting and northing as x and y coordinates to georeference them.

Defining Neighbor Relationships

The analysis of polygonal attributes is complemented by knowledge of a neighborhood structure. If measures for neighboring counties are known to be correlated, we can address this dependence when modeling an attribute of interest by other attributes. In the SIDS data, two counties are considered neighbors if their county seats are within 30 miles of each other, following the convention of Cressie (1993).

S-PLUS for ArcView manages neighbor relationships by computing an S-PLUS spatial neighbor object based on the **ncco.shp** theme data. Make sure this is the active theme before proceeding.

1. *Set the map units.* Move to ArcView. Click the View window. Choose Properties from the View menu and change the map units and the distance units to decimal degrees and miles respectively.
2. *Specify a distance threshold.* Choose Spatial Neighbors from the Spatial Statistics menu. Enter 30 next to Distance—the dialog indicates that units of miles are expected.



Figure 5.2: Calculating spatial neighbor weights in ArcView.

Note: The resulting weight matrix has a different number of neighbors than the one included with S+SPATIALSTATS, because proximities in the sids dataset are based on county seats, not centroids of the counties, as we used. Increasing the distance threshold yields a more comparable weight matrix, but Ncco.neighbor is adequate here.

3. *Create spatial neighbors.* Click the Centroid to Centroid button. We assume that the county seats and the county centroids are in close proximity in this exercise. Verify that Ncco.neighbor is under S-PLUS Neighbor Name and click OK. An S-PLUS object of class `spatial.neighbor` is created.

Now we can proceed to compute spatial correlation measures, and model the spatial structure of sudden infant death rate given other covariates.

Spatial Autocorrelation

We need to check whether the process of sudden infant deaths is spatially autocorrelated. If so, then spatial models are indeed needed to analyze these data.

1. *Open the dialog.* Choose Spatial Autocorrelation from the Spatial Statistics menu. Select `sids.arc` under S-PLUS Data Frames.
2. *Specify variables.* Under Columns, click `sid` and then `births`. Corresponding entries appear under Variables for Analysis.

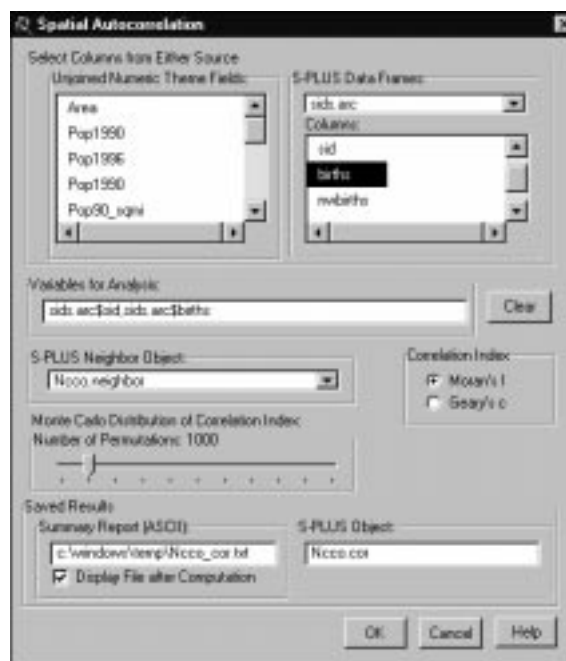


Figure 5.3: Estimating spatial autocorrelation.

3. *Specify a neighbor object.* Select `Ncco.neighbor` under S-PLUS Neighbor Object and click OK. The output below is displayed in a text editor window. The text file is in the directory indicated under Summary Report.

Note: You need to close the text editor window with the results before you can go back to ArcView.

```
*** Spatial Autocorrelation Analysis for sids.arc$sid ***

Spatial Correlation Estimate

Statistic = "moran" Sampling = "free"

Correlation = 0.1191
Variance    = 0.003835
Std. Error  = 0.06192

Normal statistic = 2.086
Normal p-value (2-sided) = 0.03695

Null Hypothesis: No spatial autocorrelation

Summary of the permutation-correlations :
      Min.  1st Qu.   Median     Mean 3rd Qu.   Max.
-0.1523 -0.05107 -0.01888 -0.01198 0.02312 0.2639

permutation p-value = 0.026

*** Spatial Autocorrelation Analysis for sids.arc$births ***

Spatial Correlation Estimate

Statistic = "moran" Sampling = "free"

Correlation = 0.1114
Variance    = 0.003835
Std. Error  = 0.06192

Normal statistic = 1.962
Normal p-value (2-sided) = 0.04982

Null Hypothesis: No spatial autocorrelation
```

Summary of the permutation-correlations :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.1593	-0.05445	-0.02006	-0.01077	0.02765	0.2433

permutation p-value = 0.034

The values for the Moran measure for spatial autocorrelation are rather high. The p-value for the significance of this coefficient is low enough that we can assume that there is enough spatial autocorrelation in the data to warrant the fitting of spatial models. One more run of the spatial autocorrelation dialog this time to compute Geary's c statistic confirms this conclusion. Geary's coefficient should lie between 0 and 1.

You can also compute the spatial correlation of arbitrary S-PLUS expressions. For example, following the reasoning in section 5.2 of the *S+SPATIALSTATS User's Manual*, a transformation of SID rate is needed to achieve constant variance in the variable of interest. This transformation is determined by the square root of the number of births times the Freeman-Tukey square root of the rates (contained in `sids.arc$sid.ft`). Enter this transformation in the dialog below Variables for Analysis.



Figure 5.4: *Entering an expression as a variable for analysis.*

This yields the following result.

```
*** Spatial Autocorrelation Analysis for sids.arc$sid.ft *
sqrt(sids.arc$births) ***
```

Spatial Correlation Estimate

```
Statistic = "moran" Sampling = "free"
```

```
Correlation = 0.2237
Variance    = 0.003835
Std. Error  = 0.06192
```

```
Normal statistic = 3.776
Normal p-value (2-sided) = 1.595e-4
```

Null Hypothesis: No spatial autocorrelation

Summary of the permutation-correlations :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.1625	-0.0522	-0.006393	-0.006191	0.03386	0.2558

permutation p-value = 0.001

This is consistent with that in the *S+SPATIALSTATS User's Manual* and indicative of high spatial autocorrelation for the transformed variable. The expected value of Moran's I is $1/(n-1)$ which in this case is about $-1/99$, so a value of 0.2237 rejects the null hypothesis of no spatial autocorrelation.

Local Spatial Association

Testing the hypothesis of no autocorrelation using the global statistics of the last section assumes stationarity of variance through space. This is not likely to be the case for most variables of interest and for the SID rates in particular. An approach that does not require the assumption of global stationarity is provided by Local Indicators of Spatial Association or LISA (Anselin, 1994, Bao and Henry, 1996). S-PLUS for ArcView provides an interface to LISA computations and display.

1. *Open the Local Spatial Association dialog.* To compute this indicator for the SIDS data, choose Local Spatial Association from the Spatial Statistics menu.
2. *Specify a field for analysis.* Select `sids.arc` below S-PLUS Data Frames and the Column `sid`.
3. *Specify a neighbor object.* Select `Ncco.neighbor` as the Spatial Neighbor object that depicts the neighborhood structure in the data, and select Geary's c statistic as the local measure of association to be used.
4. *Run local spatial association.* Click OK. This creates an S-PLUS object named `Ncco.lisa.geary`.

Since these data did not come from ArcView, you cannot automatically join the results to the active theme **ncco.shp**. You need to add a column with the county names to the results and then import the resulting table and join based on these names.

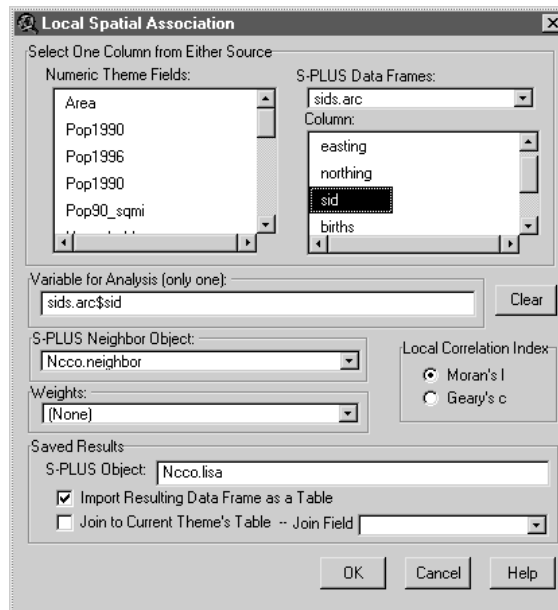


Figure 5.5: *Computing a local indicator of spatial association.*

5. *Prepare resulting data frame for import to ArcView.* To add the Name column to the data frame Ncco.lisa.geary, open up this data frame and Ncco.names by clicking their icons in the S-PLUS Object Browser. Select the column Name in

Ncco.names and by clicking in one of its rows drag it to an empty column in Ncco.lisa.geary's data window (see figure 5.6).

The figure shows two S-PLUS data windows. The window titled 'Ncco.names' contains a single column named 'Name' with 9 rows of text data. The window titled 'Ncco.local.geary' contains 6 columns of numerical data. A mouse cursor is shown dragging the 'Name' column from the first window to the second window.

	1
1	Ache
2	Alghang
3	GATY
4	Cur Buck
5	Nottingham
6	Holland
7	Carroll
8	Gates
9	Harold

	1	2	3	4	5	6
	$\bar{D}$	LocalGeary	ESP	WRI	STD	
1	-8.73	0.05	1.55	8.12	-0.60	
2	-8.90	0.25	1.78	8.81	-0.50	
3	-8.22	0.28	1.06	2.34	-0.52	
4	-8.73	0.02	1.55	8.28	-0.50	
5	-8.30	0.50	1.18	1.64	-0.47	
6	-8.94	0.30	1.01	2.82	-0.42	
7	-8.90	0.05	1.78	8.81	-0.65	
8	-8.80	0.20	1.78	4.80	-0.70	

Figure 5.6: Using drag and drop to copy the Name column from Ncco.names into Ncco.local.geary.

6. *Visualize the results.* Choose Import Table from S-PLUS from the S-PLUS menu. Enter Ncco.lisa.geary next to S-PLUS Data Frame. Fill in the dialog as shown in figure 5.7 below.

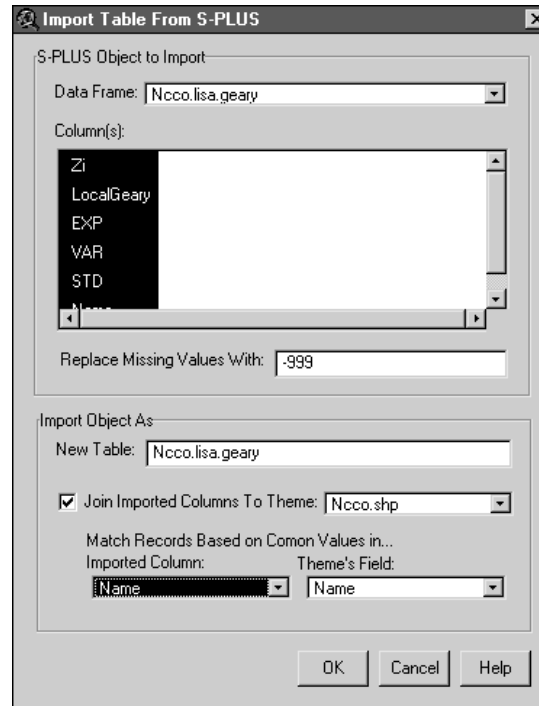


Figure 5.7: *Importing local spatial association results into ArcView.*

You can now use the Color Classification dialog to display the Local Geary's c statistics.

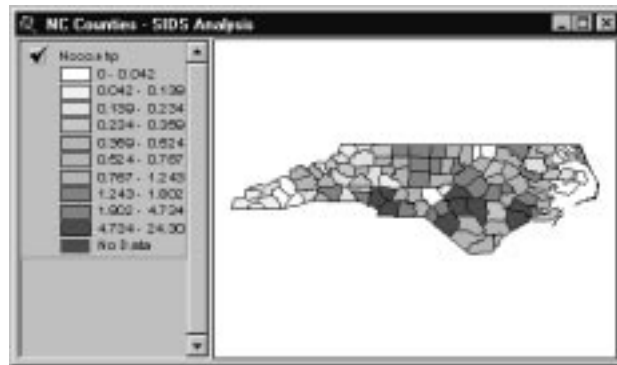


Figure 5.8: Color classification of the local correlation coefficient.

In the resulting color classification, agglomerations of both low and high values are apparent. This is characteristic of Geary's c statistic when used in this context. It detects clusters of positively correlated values—high values surrounded by high values and low values surrounded by low values.

If you run the Local Spatial Association using the transformed variable $\text{sid.ft} * \sqrt{\text{births}}$, the color classification does not pick out such striking patterns of local agglomeration. The transformed variable behaves more like a stationary process, as expected given our results using the global statistic.

Spatial Regression

To model the rate of SIDS as it is standardized by the Freeman-Tukey transformation, use spatial linear regression.

1. *Open the Spatial Linear Regression dialog.* Choose Spatial Linear Regression from the Spatial Statistics menu.
2. *Specify a response variable.* Select `sids.arc` below S-PLUS Data Frames and click `sid.ft` in the list under Columns. This will be used as the response in the formula.

3. *Specify a covariate.* To specify a covariate, click `nwbirths.ft` in the list under Columns. This field the number of non-white births, also standardized with the Freeman-Tukey transformation.
4. *Choose a covariance model.* Under Covariance Family, choose the simultaneous autoregressive covariance model (SAR). For more details about this model, and for a caveat on correcting the covariance using weights, if CAR is desired, see section 5.3 of the *S+SPATIALSTATS User's Manual*.



Figure 5.9: Spatial linear regression on the SIDS data.

5. *Fit the spatial model.* Click OK. This creates a data frame with residuals and fitted values named `Ncco.slm`.

The spatial linear regression output begins as follows.

```
*** Spatial Linear Regression Fit ***

Call:
slm(formula = sids.arc$sid.ft ~ sids.arc$nwbirths.ft,
cov.family = SAR,
  spatial.arglist = list(neighbor = Ncco.neighbor))
Residuals:
    Min       1Q   Median       3Q      Max
-2.036 -0.568  0.02794  0.4586  2.69

Coefficients:
                Value Std. Error t value Pr(>|t|)
(Intercept)  1.5428  0.2211      6.9766   0.0000
sids.arc$nwbirths.ft 0.0420  0.0062      6.7371   0.0000

Residual standard error: 0.792143 on 97 degrees of freedom

Variance-Covariance Matrix of Coefficients
                (Intercept) sids.arc$nwbirths.ft
(Intercept)    0.048901832      -0.00126093289
sids.arc$nwbirths.ft -0.001260933      0.00003884459

Correlation of Coefficient Estimates
                (Intercept) sids.arc$nwbirths.ft
(Intercept)      1.00000      -0.91488
sids.arc$nwbirths.ft -0.91488      1.00000

rho =  0.02318

Iterations =  4
Gradient norm =  4.472e-7
Log-likelihood = -206.1

Convergence:  RELATIVE FUNCTION CONVERGENCE
```

Both coefficients appear highly significant, and the covariate indeed explains some of the variability in the data.

6. *Diagnose using residuals.* The residuals from the fit are imported into ArcView and joined to the original table. This join is performed as in the example using the Local Spatial

Association results. Add the name column to the resulting data frame `Ncco.slm`, then use Import Table from S-PLUS to import and join the results in ArcView.

7. *Visualize in ArcView.* Color classification of the residuals, seen in figure 5.10, does not show any striking patterns of spatial autocorrelation left over.

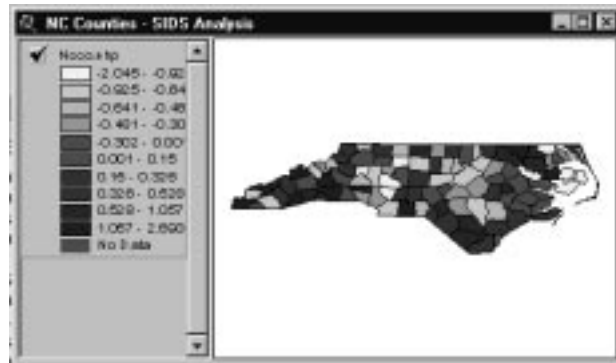


Figure 5.10: Color classification of counties by residuals of the spatial linear regression.

8. *Assess leftover spatial correlation.* The model output contains also Moran's index of spatial autocorrelation for the model residuals. For the model we just fitted these are as follows.

Spatial Correlation Estimate

Statistic = "moran" Sampling = "free"

Correlation = -0.001165

Variance = 0.003835

Std. Error = 0.06192

Normal statistic = 0.1443

Normal p-value (2-sided) = 0.8853

Null Hypothesis: No spatial autocorrelation

Summary of the permutation-correlations :

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.1899	-0.05413	-0.0176	-0.01114	0.03174	0.1805

permutation p-value = 0.409

From these results, we may conclude that the model accounts for most of the spatial autocorrelation in the data, since the residual correlation is not significant.

BIBLIOGRAPHY

- Anselin, L. (1994) Local indicators of spatial association – LISA. Research Paper 9331, Regional Research Institute, West Virginia University, Morgantown, WV.
- Bao, S. and Henry, M. S. (1996) Heterogeneity issues in local measurements of spatial association. *Geographical Systems*, 1996, **3**:1-13.
- Cressie, N. and Chan, N. H. (1989) Spatial modeling of regional variables. *Journal of American Statistical Association*, **84**:393-401.
- Cressie, N. A. C. (1993) *Statistics for Spatial Data*. Wiley, New York.

INDEX

Numerics

2D plot palette 63, 66
3D plot palette 67

A

Add Indicator Field dialog **39**
analysis of variance 71
 building a model **71**
 fitted values 73
 generating plots 73
 residuals 73
 saving results 72
 viewing printed output 73
Analysis of Variance, Fixed Effects,
 dialog **72**
Atlanta Census tracts data 9–30

B

brush and spin dialog **22**

C

clearing selection and graphics in
 ArcView 43
Color Classification dialog **40**
conditioning mode in S-PLUS 21
conditioning on a field 19

D

data frame
 removing a column 74
 sorting on a field **58**

Data window **69**
data windows in S-PLUS **13**

E

Execute S-PLUS Commands dialog **45**
executing S-PLUS commands in ArcView
 44
Export Data to S-PLUS dialog **35**
exporting
 attribute table **70**
 S-PLUS histogram to ArcView 15
 theme data to S-PLUS 34

F

Fixed Effects Analysis of Variance dialog
 72

G

Geary statistic 49, 50
glisa function 12
Graph sheet **14, 62, 65**
graphs
 importing from S-PLUS 42
Group Selected Records dialog **38**
grouping records 37

H

histogram 66
 changing the number of bars 66
 customizing 66
histograms in S-PLUS 13–15

I

- Import Point Theme dialog **37**
- Import Table From S-PLUS dialog **36**
- importing
 - data from S-PLUS 36
 - graphs from S-PLUS 42
 - point theme from S-PLUS 37
 - table from S-PLUS 75
- indicator field 24
- installing 2
 - over S-PLUS 4.0 2
- interactive brush and spin plot
 - printing 23
- interactive brush and spin plots **22**

L

- legend editor 26
- linear regression **60, 64**
 - generating plots from 65
 - generating plots with 62
 - model in ArcView 43
 - plot of a three dimensional fit 67
 - residuals 63
 - saving results 62
 - saving results of 64
 - specifying a model 61
 - versus spatial linear regression 25
 - viewing the results of 62
 - visualizing the fit in three dimensions 67
- Linear Regression dialog **24, 44, 62**
- local spatial association 50
- Local Spatial Association dialog **51**

M

- menu
 - S-PLUS 34
- model errors, *see* linear regression, residuals
- Moran statistic 49, 50

O

- Object Browser **12, 56**
 - displaying object details 57
 - filtering with 12, **57**

- folders in 63
- opening 56
- right pane property dialog **57**
- viewing a data frame **58**
- viewing classes 57
- viewing data frames with 13
- viewing databases 57

R

- Report window **59, 62, 73**
 - formatting **59**
- residuals, *see* linear regression, residuals
- Rotate 67
- running S-PLUS for ArcView 32

S

- scatter plot **63**
 - matrix **18**
 - three-dimensional **22**
 - with loess curve fit **19**
- scatter plot matrix **18**
- spatial autocorrelation 49
- Spatial Autocorrelation dialog **49**
- spatial bar chart 41
- Spatial Bar/Pie Chart dialog **41**
- spatial barchart **27, 28**
- spatial linear regression 52
 - building a model formula 52
- Spatial Linear Regression dialog **52**
- spatial neighbor object 51
- spatial neighbors 46–49
- Spatial Neighbors dialog **46**
- spatial pie chart 41
- spatial statistics
 - menu 46
 - module 46
- spatial statistics menu **2, 46**
- spatial weights 46–49
 - adjusted neighbor 47
 - border-to-border distance 48
 - centroid-to-centroid distance 48
 - distance units 48
 - first order neighbor 46
 - higher order neighbor 48
- spline plot **69**
- S-Plus for ArcView
 - running 32

- tips on learning 4–5
- S-PLUS menu **2, 34**
- Subset dialog **60**
- subsetting the rows of a data frame 60
- summary statistics
 - generating 58
- Summary Statistics dialog **58**
- sweeping data in a plot **23**

T

- technical support 6
- Transform dialog **64**
- transforming a field 64
- typographic conventions 5

U

- U. S. States data 42

V

- viewing statistical output **59**

W

- weighting data 24
- weights
 - specifying in a linear regression model 43

